EFFECTOR PROTEINS OF RAPAMYCIN

Related Applications

This application is a continuation-in-part of co-pending patent application Serial No. 08/384,524, filed February 13, 1995, which is a continuation-in-part of patent application Serial No. 08/312,023, filed September 26, 1994, now abandoned, which is a continuation-in-part of patent application Serial No. 08/207,975, filed March 8, 1994, now abandoned.

This invention concerns effector proteins of Rapamycin. More particularly, this invention concerns novel Rapamycin-FKBP12 binding proteins of mammalian origin for identification, design and synthesis of immunomodulatory, anti-restenosis or anti-tumor agents.

BACKGROUND OF THE INVENTION

Rapamycin is a macrolide antibiotic produced by *Streptomyces hygroscopicus* which was first characterized via its properties as an antifungal agent. It adversely affects the growth of fungi such as *Candida albicans* and *Microsporum gypseum*. Rapamycin, its preparation and its antibiotic activity were described in U.S. Patent No. 3,929,992, issued December 30, 1975 to Surendra Sehgal et al. In 1977 Martel, R. R. et al. reported on immunosuppressive properties of rapamycin against experimental allergic encephalitis and adjuvant arthritis in the Canadian Journal of Physiological Pharmacology, 55, 48-51 (1977). In 1989, Calne, R. Y. et al. in Lancet, 1989, no. 2, p. 227 and Morris, R. E. and Meiser, B. M. in Medicinal Science Research, 1989, No. 17, P. 609-10, separately reported on the effectiveness of rapamycin in inhibiting rejection *in vivo* in allograft transplantation. Numerous articles have followed describing the immunosuppressive and rejection inhibiting properties of rapamycin, and clinical investigation has begun for the use of rapamycin in inhibiting rejection in transplantation in man.

Rapamycin alone (U.S. Patent 4,885,171) or in combination with picibanil (U.S. Patent 4,401,653) has been shown to have antitumor activity. R. R. Martel et al. [Can. J. Physiol. Pharmacol. 55, 48 (1977)] disclosed that rapamycin is effective in the experimental allergic encephalomyelitis model, a model for multiple sclerosis; in the

adjuvant arthritis model, a model for rheumatoid arthritis; and effectively inhibited the formation of IgE-like antibodies.

The immunosuppressive effects of rapamycin have been disclosed in FASEB 3, 3411 (1989). Cyclosporin A and FK-506, other macrocyclic molecules, also have been shown to be effective as immunosuppressive agents, therefore useful in preventing transplant rejection [FASEB 3, 3411 (1989); FASEB 3, 5256 (1989); R. Y. Calne et al., Lancet 1183 (1978); and U.S. Patent 5,100,899].

Rapamycin has also been shown to be useful in preventing or treating systemic lupus erythematosus [U.S. Patent 5,078,999], pulmonary inflammation [U.S. Patent 5,080,899], insulin dependent diabetes mellitus [Fifth Int. Conf. Inflamm. Res. Assoc. 121 (Abstract), (1990)], and smooth muscle cell proliferation and intimal thickening following vascular injury [Morris, R. J. Heart Lung Transplant 11 (pt. 2): 197 (1992)].

Mono- and diacylated derivatives of rapamycin (esterified at the 28 and 43 positions) have been shown to be useful as antifungal agents (U.S. Patent 4,316,885) and used to make water soluble prodrugs of rapamycin (U.S. Patent 4,650,803). Recently, the numbering convention for rapamycin has been changed; therefore according to Chemical Abstracts nomenclature, the esters described above would be at the 31- and 42- positions. U.S. Patent 5,118,678 discloses carbamates of rapamycin that are useful as immunosuppressive, anti-inflammatory, antifungal, and antitumor agents. U.S. Patent 5,100,883 discloses fluorinated esters of rapamycin. U.S. Patent 5,118,677 discloses amide esters of rapamycin. U.S. Patent 5,130,307 discloses aminoesters of rapamycin. U.S. Patent 5,117,203 discloses sulfonates and sulfamates of rapamycin. U.S. Patent 5,194,447 discloses sulfonylcarbamates of rapamycin.

U.S. Patent No. 5,100,899 (Calne) discloses methods of inhibiting transplant rejection in mammals using rapamycin and derivatives and prodrugs thereof. Other chemotherapeutic agents listed for use with rapamycin are azathioprine, corticosteroids, cyclosporin (and cyclosporin A), and FK-506, or any combination thereof.

Rapamycin produces immunosuppressive effects by blocking intracellular signal transduction. Rapamycin appears to interfere with a calcium independent

signalling cascade in T cells and mast cells [Schreiber et al. (1992) Tetrahedron 48:2545-2558]. Rapamycin has been shown to bind to certain immunophilins which are members of the FK-506 binding proteins (FKBP) family. In particular, Rapamycin has been shown to bind to the binding proteins, FKBP12, FKBP13, FKBP25 [Galat A. et al., (1992) Biochemistry 31(8);2427-2437 and Ferrera A, et al., (1992) Gene 113(1):125-127; Armistead and Harding, Ann. Reports in Med. Chem. 28:207-215, 1993], and FKBP52 [WO 93/07269]

Rapamycin is able to inhibit mitogen-induced T cell and B cell proliferation as well as proliferation induced by several cytokines, including IL-2, IL-3, IL-4 and IL-6 (reviewed by Sehgal et al., Med. Research Rev.14: 1-22, 1994). It can also inhibit antibody production. Rapamycin has been shown to block the cytokine-induced activation of $p70^{S6}$ kinase which appears to correlate with Rapamycin's ability to decrease protein synthesis accompanying cell cycle progression (Calvo et al., Proc. Natl. Acad. Sci. USA, 89:7571-7575,1992; Chung et al., Cell 69:1227-1236, 1992; Kuo et al., Nature 358:70-73,1992; Price et al., Science 257:973-977, 1992). It also inhibits the activation of cdk2/cyclin E complex (Flanagan et al., Ann. N.Y.Acad. Sci, in press; Flanagan et al, Mol. Cell biol., in press; Flanagan et al., J.Cell Biochem. 17A:292, 1993). Rapamycin's effects are not mediated by direct binding to $p70^{s6}$ kinase and cdk2/cyclin E, but by action of the Rapamycin-FKBP complex on upstream component(s) which regulate the activation status of the kinases.

It is generally accepted that the action of immunosuppressive drugs, such as Rapamycin, cyclosporine and FK506, is dependent upon the formation of a complex with their respective intracellular receptor proteins called immunophilins. While the binding of these immunosuppressants with their respective immunophilins inhibits the cis-trans peptidyl prolyl isomerase (PPIase) activity of immunophilins, PPIase inhibition is not sufficient to mediate the immunosuppressive activity (reviewed in Armistead and Harding, Annual Reports in Med. Chem, 28:207-215:1993). Two rapamycin analogs which are Diels Alder adducts, one with 4-phenyl-1,2,4-triazoline-3,5-dione, and the second with 4-methyl-1,2,4-triazoline-3,5-dione, bind to FKBP, inhibited its PPIase activity, yet they did not exhibit any detectable immunosuppressive activity. The phenyl-triazolinedione Diels Alder adduct at high molar excess has been shown to competitively inhibit rapamycin's effect on DNA synthesis in mitogen-

stimulated murine thymocyte proliferation (Ocain et al., Biochem. Biophys. Res. Commun. 192:1340, 1993). Recent evidence suggests that the binary immunophilin-drug complex such as cyclophilin-cyclosporin A and FKBP-FK506 gains a new function that enables it to block signal transduction by acting on specific target proteins. The molecular target of both cyclophilin-cyclosporin A and FKBP-FK506 complexes such as has been identified as the Ca^{+2}/calmodulin dependent serine/threonine phosphatase calcineurin (J. Liu et al, Cell 66, 807, 1991; J. Liu et al, Biochemistry 31, 3896, 1992; W.M. Flanagan, et al., Nature 352, 803, 1992; McCaffrey et al., J. Biol. Chem. 268, 3747, 1993; McCaffrey et al., Science 262:750, 1993).

Rapamycin's antifungal and immunosuppressive activities are mediated via a complex consisting of Rapamycin, a member of the FK506 binding protein (FKBP) family and at least one additional third protein, called the target of Rapamycin (TOR). The family of FKBPs is reviewed by Armistead and Harding (Annual Reports in Med. Chem, 28:207-215:1993). The relevant FKBP molecule in Rapamycin's antifungal activity has been shown to be FKBP12 (Heitman et al., Science 253:905-909:1993). In mammalian cells, the relevant FKBPs are being investigated. Although two TOR proteins (TOR1 and TOR2) have been identified in yeast (Kunz et al., Cell 73:585-596:1993), the target of Rapamycin in human cells remains elusive. The carboxy terminus of yeast TOR2 has been shown to exhibit 20% identity with two proteins, the p110 subunit of phosphatidylinositol 3-kinase and VPS34, a yeast vacuolar sorting protein also shown to have PI 3K activity. However, J. Blenis et al. (AAI meeting, May, 1993) have reported that Rapamycin-FKBP12 complex does not directly mediate its effects on PDGF stimulated cells via the p110, p85 PI 3K complex.

DESCRIPTION OF THE INVENTION

This invention concerns isolated, cloned and expressed proteins which bind to a complex of GST-FKBP12-Rapamycin. These proteins are isolated from membrane preparations of Molt 4 T cell leukemia. The sizes of the four novel proteins are estimated by PAGE migration to be 125±12 kilodaltons (kDa), 148±14 kDa, 208± 15 kDa and 210±20kDa and will be referred to herein and in the claims that follow, as the 125 kDa, 148 kDa, 208 kDa, and 210 kDa, respectively. The four proteins may also be referred to herein as effector proteins.

The proteins of this invention can be used in screening assays, such as enzyme inhibitor assays and binding assays, to identify endogenous complexes and ligands and novel exogenous compounds (like Rapamycin) which modulate their functions. They can also be used in assays to identify compounds with therapeutic benefit for restenosis, immunomodulation and as antitumor agents. Cloning the proteins of this invention does not only allow the production of large quantities of the proteins, it also provides a basis for the development of related anti-sense therapeutics. The use of cDNA clones to generate anti-sense therapeutics with immunomodulatory activity (for use against transplantation rejection, graft versus host disease, autoimmune diseases such as lupus, myasthenia gravis, multiple sclerosis, rheumatoid arthritis, type I diabetes, and diseases of inflammation such as psoriasis, dermitis, eczema, seborrhea, inflammatory bowel disease, pulmonary inflammation, asthma, and eye uveitis), antirestenosis and anti-tumor activity is included within the scope of this invention.

The proteins of the present invention can be isolated from mammalian cells, such as cells of the T cell leukemia cell line, Molt 4 (ATCC 1582, American Type Cell Culture, 12301 Parklawn Drive, Rockville, MD, USA, 20852), the B cell lymphoma, BJAB, or normal human T cells. These mammalian cells can be lysed in a buffer containing protease inhibitors and reducing agent (2-ME), such as hypotonic buffer A (100 mM HEPES, pH 7.5, 20 mM KCl, 1 mM EDTA, 0.4 mM PMSF and 2 mM beta mercaptoethanol (2-ME)). The cell nuclei and unbroken cells are cleared by centrifugation at a temperature which minimizes protein degradation. The membrane fraction of the cells can then be concentrated or pelleted by ultracentrifugation at 100,000 g. Detergent solubilization of the membrane pellet is carried out in a detergent containing buffer such as buffer B (50 mM Tris, pH 7.2, 100 mM NaCl, 20 mM KCl, 0.2 mM PMSF, 1 mM 2-ME, 2 mM $CaCl_2$, 2 mM $MgCl_2$, 5 µg/ml aprotinin, leupeptin, pepstatin A and antipain), containing CHAPSO (3-[(3-cholamido-propyl)dimethylammonio]-1-propane sulfonate; 12 mM) or Triton X100 (polyethylene glycol 4-isooctylphenyl ether). The solubilized membrane proteins can then be separated from the debris by 100,000g ultracentrifugation at a temperature which minimizes protein degradation. The supernatant containing solubilized membrane proteins is then preabsorbed with an affinity resin, such as glutathione resin, in the presence of protease inhibitors at a temperature which minimizes protein degradation..

After centrifugation to remove the resin from the supernatant, the supernatant is then incubated with complexed Rapamycin or Rapamycin analog to FKBP, such as GST-FKBP12--Rapamycin at a temperature which minimizes protein degradation. The mixture of solubilized membrane proteins, incubated with complexed Rapamycin or Rapamycin analog to FKBP, such as GST-FKBP12--Rapamycin, can then be incubated with the affinity resin to bind the complexes of rapamycin or rapamycin analog, FKBP fusion protein and binding proteins at a temperature which minimizes protein degradation. After most non-specific proteins are rinsed away using a detergent containing buffer, such as Buffer C (50 mM Tris, pH 7.2, 100 mM NaCl, 20 mM KCl, 0.2 mM PMSF, 1 mM 2-ME or 10 mM dithiothreitol, 0-5 mM $CaCl_2$, 0-5 mM $MgCl_2$, 5 µg/ml aprotinin, leupeptin, pepstatin A and antipain and 0.1% Triton X100) (Polyethylene glycol 4-isooctyl phenyl ether), the proteins are eluted from the resin under denaturing conditions, such as a buffer containing sufficient detergent to dissociate it from resin (e.g. Laemli buffer with or without glycerol or dye, as described by Laemli, Nature 227:680, 1970), or non-denaturing conditions such as a buffer containing an appropriate eluting compound for the affinity column, such as 5 mM glutathione. The proteins can then be separated by size using SDS polyacrylamide gel electrophoresis (SDS-PAGE).

The present invention also includes the genomic DNA sequences for the abovementioned proteins, as well as the cDNA and anti-sense RNA and DNA sequences which correspond to the genes for the abovementioned proteins. The present invention further includes the proteins of other mammalian species which are homologous or equivalent at least in function to the abovementioned proteins, as well as the DNA gene sequences for the homologous or equivalent proteins and the cDNA and anti-sense RNA and DNA sequences which correspond to the genes for the homologous or equivalent proteins.

For the purposes of this disclosure and the claims that follow, equivalents of the proteins of this invention are considered to be proteins, protein fragments and/or truncated forms with substantially similar, but not identical, amino acid sequences to the proteins mentioned above, the equivalents exhibiting rapamycin-FKBP complex binding characteristics and function similar to the proteins mentioned above. Therefore, in this specification and the claims below, references to the 125 kDa, 148

kDa, 208 kDa, and 210 kDa proteins of this invention are also to be understood to indicate and encompass homologous or equivalent proteins, as well as fragmented and/or truncated forms with substantially similar, but not identical, amino acid sequences of the 125 kDa, 148 kDa, 208 kDa, and 210 kDa proteins mentioned above.

These proteins or protein homologues or equivalents can be generated by similar isolation procedures from different cell types and/or by recombinant DNA methods and may be modified by techniques including site directed mutagenesis. For example, the genes of this invention can be engineered to express one or all of the proteins as a fusion protein with the fusion partner giving an advantage in isolation (e.g. HIS oligomer, immunoglobulin Fc, glutathione S-transferase, FLAG etc). Mutations or truncations which result in a soluble form can also be generated by site directed mutagenesis and would give advantages in isolation.

This invention further includes oligopeptide fragments, truncated forms and protein fragments that retain binding affinity yet have less than the active protein's amino acid sequences. This invention also includes monoclonal and polyclonal antibodies specific for the proteins and their uses. Such uses include methods for screening for novel agents for immunomodulation and/or anti-tumor activity and methods of measuring the parent compound and/or metabolites in biological samples obtained from individuals taking immunosuppressive drugs. The use of the cDNA clone to generate anti-sense therapeutics (Milligan et al, J. Med. Chem. 36:1923-1936, 1993) with immunomodulatory activity (transplantation rejection, graft versus host disease, autoimmune diseases such as lupus, myasthenia gravis, multiple sclerosis, rheumatoid arthritis, type I diabetes, and diseases of inflammmation such as psoriasis, dermitis, eczema, seborrhea, inflammatory bowel disease, pulmonary inflammation, asthma, and eye uveitis), and anti-tumor activity is also included in the present invention.

The proteins of this invention can also be made by recombinant DNA techniques familiar to those skilled in the art. That is, the gene of the protein in question can be cloned by obtaining a partial amino acid sequence by digestion of the protein with a protease, such as Lysine C, and isolating the resulting protein fragments by microbore HPLC, followed by fragment sequencing (Matsudaira in A Practical

Guide to Protein and Peptide Purification for Microsequencing, Academic Press (San Diego, CA, 1989)). The determined sequence can then be used to make oligonucleotide probes which can be used to screen a human cDNA library directly or generate probes by polymerase chain reaction. The library can be generated from human T cells or the cell lines, Molt 4, Jurkat, or other etc. to obtain clones. These clones can be used to identify additional clones containing additional sequences until the protein's full gene, i.e. complete open reading frame, is cloned.

It is known in the art that some proteins can be encoded by an open reading frame which is longer than initially predicted by the size of the protein. These proteins may represent cleavage products of the precursor protein translated from the complete open reading frame (eg. IL-1 beta) or proteins translated using a downstream start codon (eg. Hepaptitis B surface antigen). In view of this knowledge, it is understood that the term cDNA as used herein and in the claims below refers to cDNA for the gene's complete open reading frame or any portions thereof which may code for a protein of this invention or the protein's fragments, together or separate, or truncated forms, as previously discussed.

In a complementary strategy, the gene(s) for the proteins of this invention may be identified by interactive yeast cloning techniques using FKBP12:RAPA as a trap for cloning. These strategies can also be combined to quicken the identification of the clones.

The relevant cDNA clone encoding the gene for any of the four proteins can also be expressed in E. coli, yeast, or baculovirus infected cells or mammalian cells using state of the art expression vectors. Isolation can be performed as above or the cDNA can be made as a fusion protein with the fusion partner giving an advantage in isolation (e.g. HIS oligomer, immunoglobulin Fc, glutathione S-transferase, etc). Mutations which result in a soluble form can also be generated by site directed mutagenesis and would give advantages in isolation.

The uses of such cDNA clones include production of recombinant proteins. Further, such recombinant proteins, or the corresponding natural proteins isolated from mammalian cells, or fragments thereof (including peptide oligomers) are useful in

generation of antibodies to these proteins. Briefly, monoclonal or polyclonal antibodies are induced by immunization with recombinant proteins, or the corresponding natural proteins isolated from mammalian cells, or fragments thereof (including peptide oligomers conjugated to a carrier protein (e.g. keyhole limpet hemocyanin or bovine serum albumin)) of animals using state of the art techniques. The antibodies can be used in the purification process of the natural proteins isolated from mammalian cells or recombinant proteins from E. coli, yeast, or baculovirus infected cells or mammalian cells, or cell products.

The uses of such cDNA clones include production of recombinant proteins. Further, such recombinant proteins, or the corresponding natural proteins isolated from mammalian cells, are useful in methods of screening for novel agents such as synthetic compounds, natural products, exogenous or endogenous substrates for immunomodulation and/or antitumor activity. The natural products which may be screened may include, but are not limited to, cell lysates, cell supernatants, plant extracts and the natural broths of fungi or bacteria. As an example of a competitive binding assay, one of these proteins attached to a matrix (either covalently or noncovalently) can be incubated with a buffer containing the compounds, natural products, cell lysates or cell supernatants and a labeled rapamycin:FKBP complex. The ability of the compound, natural products, exogenous or endogenous substrates to competitively inhibit the binding of the complex or specific antibody can be assessed. Examples of methods for labeling the complex include radiolabeling, fluorescent or chemiluminescent tags, fusion proteins with FKBP such as luciferase, and conjugation to enzymes such as horse radish peroxidase, alkaline phosphatase, acetylcholine esterase (ACHE), etc. As an example of an enzymatic assay, the proteins are incubated in the presence or absence of novel agents such as synthetic compounds, natural products, exogenous or endogenous substrates with substrate and the enzymatic activity of the protein can be assessed. Methods of measuring the parent compound and/or metabolites in biological samples obtained from individuals taking immunosuppressive drugs can also be assessed using these proteins.

This invention includes a method for identifying substances which may be useful as immunomodulatory agents or anti-tumor agents, the method utilizing the following steps:

a) combining the substance to be tested with one of the four mammalian proteins (125 kDa, 148kDa, 208 kDa or 210 kDa) of this invention, with the protein being bound to a solid support:

b) maintaining the substance to be tested and the protein bound to the solid support of step (a) under conditions appropriate for binding of the substance to be tested with the protein, and

c) determining whether binding of the substance to be tested occurred in step (b).

This invention also includes a method for identifying substances which may be useful as immunomodulatory or anti-tumor agents which involves the following steps:

a) combining a substance to be tested with one of the mammalian proteins of this invention, the protein being bound to a solid support:

b) maintaining the substance to be tested and the protein bound to the solid support of step (a) under conditions appropriate for binding of the substance to be tested with the protein, and

c) determining whether the presence of the substance to be tested modulated the activity of the mammalian protein.

This invention further includes a method for detecting, in a biological sample, rapamycin, rapamycin analogs or rapamycin metabolites which, when complexed with a FKBP, bind to one of the four listed proteins of this invention, the method comprising the steps of:

a) combining the biological sample with a FKBP to form a first mixture containing, if rapamycin, rapamycin analogs or rapamycin metabolites are present in the biological sample, a rapamycin:FKBP complexes, rapamycin analog:FKBP complexes, or rapamycin metabolite:FKBP complexes;

b) creating a second mixture by adding the first mixture to one of the proteins of this invention, the protein bound to a solid support ;

c) maintaining the second mixture of step (b) under conditions appropriate for binding the rapamycin:FKBP complexes, rapamycin analog:FKBP complexes, or rapamycin metabolite:FKBP complexes, if present, to the protein of this invention; and

d) determining whether binding of the rapamycin:FKBP complexes, rapamycin analog:FKBP complexes, or rapamycin metabolite:FKBP complexes and the protein occurred in step (c).

Also included in this invention is the use of the cDNA clones to generate anti-sense therapeutics. This can be accomplished by using state of the art techniques, such as those described in Milligan et al, J. Med. Chem. 36:14:1924-1936. For the purposes of this disclosure and the claims that follow, antisense RNA and DNA are understood to include those RNA and DNA strands derived from a cDNA clone which encodes for one of the four proteins (125 kDa, 148 kDa, 208 kDa or 210 kDa) of the present invention which have a native backbone or those which utilize a modified backbone. Such modifications of the RNA and DNA backbones are described in Milligan et al, J. Med. Chem. 36:14:1924-1936. The antisense compounds created by the state of the art techniques recently described (Milligan et al, J. Med. Chem. 36:14:1924-1936) can be useful in modulating the immune response and thus useful in the treatment or inhibition of transplantation rejection such as kidney, heart, liver, lung, bone marrow, pancreas (islet cells), cornea, small bowel, and skin allografts, and heart valve xenografts; in the treatment or inhibition of autoimmune diseases such as lupus, rheumatoid arthritis, diabetes mellitus, myasthenia gravis, and multiple sclerosis; and diseases of inflammation such as psoriasis, dermatitis, eczema, seborrhea, inflammatory bowel disease, and eye uveitis. The antisense molecules of this invention can have antitumor, antifungal activities, and antiproliferative activities. The compounds of this invention therefore can be also useful in treating solid tumors, adult T-cell leukemia/lymphoma, fungal infections, and hyperproliferative vascular diseases such as restenosis and atherosclerosis. Thus, the present invention also comprises

methods for treating the abovementioned maladies and conditions in mammals, preferably in humans. The method comprises administering to a mammal in need thereof an effective amount of the relevant antisense therapeutic agent of this invention.

When administered for the treatment or inhibition of the above disease states, the antisense molecules of this invention can be administered to a mammal orally, parenterally, intranasally, intrabronchially, transdermally, topically, intravaginally, or rectally.

It is contemplated that when the antisense molecules of this invention are used as an immunosuppressive or antiinflammatory agent, they can be administered in conjunction with one or more other immunoregulatory agents. Such other immunoregulatory agents include, but are not limited to azathioprine, corticosteroids, such as prednisone and methylprednisolone, cyclophosphamide, rapamycin, cyclosporin A, FK-506, OKT-3, and ATG. By combining the complexes of this invention with such other drugs or agents for inducing immunosuppression or treating inflammatory conditions, the lesser amounts of each of the agents are required to achieve the desired effect. The basis for such combination therapy was established by Stepkowski whose results showed that the use of a combination of rapamycin and cyclosporin A at subtherapeutic doses significantly prolonged heart allograft survival time. [Transplantation Proc. 23: 507 (1991)].

Treatment with these antisense compounds will generally be initiated with small dosages less than the optimum dose of the compound. Thereafter the dosage is increased until the optimum effect under the circumstances is reached. Precise dosages will be determined by the administering physician based on experience with the individual subject treated. In general, the antisense compounds of this invention are most desirably administered at a concentration that will afford effective results without causing any harmful or deleterious side effects.

In light of the therapeutic value of the abovementioned antisense compounds, this invention also includes pharmaceutical compositions containing the antisense RNA and antisense DNA compounds derived from cDNA clones which encode for the 125 kDa, 148 kDa, 208 kDa and 210 kDa proteins of this invention.

This invention also comprises the following process for isolating the proteins of this invention, as well as the proteins isolated therefrom:

A process for isolating proteins from mammalian cells, the process comprising the steps of:

1. The mammalian cells of interest are grown and harvested. As mentioned previously, the cells may be of T cell origin (e.g. T cell lymphomas, leukemias, normal human T cells), B cell origin (e.g. EBV transformed B cells, normal human B cells), mast cells, or other cell sources sensitive to rapamycin. The cells may be processed shortly after harvesting or may be stored frozen, such as in pellets, prior to processing. The cells which are kept frozen may be stored in a dry ice and ethanol bath, stored frozen at -70-80° C until use. This step of growing and harvesting the cells of interest may be seen as the first step of this process or as merely preparatory for the present process.

2. Cells are lysed in a buffer containing a buffering agent (e.g.HEPES, Tris, pH 7.5), low salt (e.g.10 -50 mM NaCl or KCl), chelating agent (e.g. 1-2 mM EDTA), protease inhibitors (e.g.0.4 mM PMSF) and a reducing agent (e.g. 2 mM 2-ME or 1-20 mM Dithiothreitol) at a temperature which minimizes protein degradation (e.g. 4 °C). It should be understood that the mammalian cells may be treated in any manner capable of producing cell lysis, including sonic lysis and douncing.

3. Unbroken cells and cell nuclei are precleared from lysates by centrifugation at a temperature which minimizes protein degradation (e.g. 4 °C). Centrifugation at, for example, 1600g for 10 minutes has been found sufficient to preclear the unbroken cells and cell nuclei from the lysates. This step, while not mandatory, provides a clearer preparation for the steps that follow.

4. The membrane fraction in the precleared lysate is then concentrated, such as by ultracentrifugation. An example of this concentration would be ultracentrifugation at 100,000 g for 1-1.5 hours.

5. The membrane proteins (e.g. transmembrane, integral and membrane associated proteins) are then solubilized. This may be accomplished by incubating the pellet of Step 4 in a buffer containing a detergent which solubilizes the proteins without detrimentally denaturing them, a buffering agent (e.g. 20-50 mM Tris or HEPES, pH 7.2), salt (e.g. 100 - 200 mM NaCl + 20 mM KCl), reducing agent (e.g. 1-2 mM 2-ME or 1 - 20 mM dithiothreitol), protease inhibitors (e.g. 0.2 mM PMSF, 5 µg/ml aprotinin, leupeptin, pepstatin A and antipain), divalent cations (e.g. 0-5 mM $CaCl_2$, 0-5 mM $MgCl_2$) at a temperature which minimizes protein degradation (e.g. 4° C) . Examples of detergents useful in this step are CHAPSO (3-[(3-cholamidopropyl)dimethylammonio]-1-propane sulfonate) or Triton X100 (polyethylene glycol 4-isooctylphenyl ether). After this step, the mixture contains solubilized membrane proteins and non-solubilized cellular debris.

6. The solubilized membrane proteins are separated from the non-solubilized cellular debris, such as by ultracentrifugation (eg 100,000g for 1-1.5 hours) at a temperature which minimizes protein degradation (e.g. 4 °C).

7. The supernatant containing solubilized membrane proteins is incubated with an affinity resin in a buffer containing a buffering agent (e.g.20-50 mM Tris or HEPES, pH 7.2), salt (e.g. 100 - 200 mM NaCl + 20 mM KCl), reducing agent (e.g. 1-2 mM 2-ME or 10 - 20 mM dithiothreitol), protease inhibitors (e.g. 0.2 mM PMSF, 5 µg/ml aprotinin, leupeptin, pepstatin A and antipain), divalent cations (e.g. 0-5 mM $CaCl_2$, 0-5 mM $MgCl_2$) at a temperature and time which allows the absorption of the proteins which bind to affinity resin directly, and minimizes protein degradation (e.g. 4 °C).

8. The resin is then removed from the supernatant by centrifugation at a temperature which minimizes protein degradation (e.g. 4 °C).

9. The supernatant is then incubated with Rapamycin or Rapamycin analog (IC50 in LAF < 500nM) complexed to fusion protein of FKBP12 +protein which enhances the isolation of the desired effector protein and through which the fusion protein binds to an affinity resin or affinity column, such as GST-FKBP12,

Histidine oligomer -FKBP12, biotin-FKBP12, etc., in a buffer containing a buffering agent (e.g. 20-50 mM Tris or HEPES, pH 7.2), salt (e.g. 100 - 200 mM NaCl + 20 mM KCl), reducing agent (e.g. 1-2 mM 2-ME or 1 - 20 mM dithiothreitol), protease inhibitors (e.g. 0.2 mM PMSF, 5 µg/ml aprotinin, leupeptin, pepstatin A and antipain), divalent cations (e.g. 0-5 mM $CaCl_2$, 0-5 mM $MgCl_2$) at a temperature and for a time which allows binding of the effector proteins to the fusion FKBP protein:Rapamycin or analog complexes and minimizes protein degradation (e.g. 4 °C and 1-2 hours).

10. The mixture of Step 9 containing the effector proteins and fusion FKBP protein:Rapamycin complexes is incubated with an affinity resin at a temperature and for a time which allows binding of the complexes of the effector proteins and fusion FKBP protein:Rapamycin or analog to the affinity resin and minimizes protein degradation (e.g. 4 °C and 0.5-2 hours).

11. Most non-specific proteins are rinsed away from the resin using a buffer which dissociates binding of non-specific proteins but not the complex between the desired proteins and RAPA-FKBP, such as a buffer containing a buffering agent (e.g.20-50 mM Tris or HEPES, pH 7.2), salts (e.g. 100 - 1000 mM NaCl, KCl), reducing agent (e.g. 1-2 mM 2-ME or 10 - 20 mM dithiothreitol), protease inhibitors (e.g. 0.2 mM PMSF, 5 µg/ml aprotinin, leupeptin, pepstatin A and antipain), divalent cations (e.g. 0-5 mM $CaCl_2$, 0-5 mM $MgCl_2$) and detergent which dissociates binding of non-specific proteins but not the complex between the four proteins and RAPA-fusion FKBP protein such as Triton X100 (Polyethylene glycol 4-isooctyl phenyl ether).

12. The effector proteins and the fusion FKBP protein:Rapamycin complexes are eluted from the resin using an appropriate buffer, such as a buffer containing sufficient detergent to dissociate it from resin (e.g. Laemli buffer with or without glycerol or dye, Laemli, Nature 227:680, 1970), or an appropriate eluting compound for the affinity column, such as glutathione, histidine.

13. The effector proteins can then be separated by size. This may be accomplished in any manner which separates the proteins by size, including, but not

limited to, polyacrylamide gel electrophoresis and size exclusion column chromatography.

It might also be useful to compare the proteins isolated by a control procedure, that is a procedure which substitutes buffer for the rapamycin or rapamycin analog with an IC_{50} in LAF < 500 nM in step 8, can be used to more easily distinguish proteins which bind to the rapamycin:FKBP complex.

The proteins of this invention can also be made by recombinant DNA techniques familiar to those skilled in the art. That is, the gene of the protein in question can be cloned by obtaining a partial amino acid sequence by digestion of the protein with an appropriate endopeptidase, such as Lysine C, and isolating the resulting protein fragments by microbore HPLC, followed by fragment sequencing (Matsudaira in A Practical Guide to Protein and Peptide Purification for Microsequencing, Academic Press, San Diego, CA 1989). The determined sequence can then be used to make oligonucleotide probes which can be used to screen a human cDNA library, such as those for human T cells, Molt 4, Jurkat, etc, to obtain clones.(Sambrook, Fritsch, and Maniatas, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, 1989) These clones can be used to identify additional clones containing additional sequences until the protein's full gene is cloned (Sambrook, Fritsch, and Maniatas, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, 1989). In a complementary strategy, the gene(s) may be identified by interactive yeast cloning techniques using FKBP12:RAPA as a trap for cloning (Chien et al., Proc. Natl. Acad. Sci. 88: 9578-9582, 1991). These strategies can also be combined to quicken the identification of the clones.

The relevant cDNA clone can also be expressed in E.coli, yeast, or baculovirus infected cells or mammalian cells using state of the art expression vectors. Isolation can be performed as above or the cDNA can be made as a fusion protein with the fusion partner giving an advantage in isolation (e.g. HIS oligomer, immunoglobulin Fc, glutathione S-transferase, etc). Mutations which result in a soluble form can also be generated by site directed mutagenesis and would give advantages in isolation.

Homologs in the mouse, rat, monkey, dog and other mammalian species can be obtained using similar procedures. In addition, upon isolation of the human clone of the proteins, the clone can be used to screen for homologs in other mammalian species. These homologs can also be used to develop binding assays and to set up high through put screening assays for compounds, endogenous ligands, exogenous ligands with immunomodulatory activity.

Compounds, endogenous ligands and exogenous ligands having such immunomodulatory activity would can be useful in modulating the immune response and thus useful in the treatment or inhibition of transplantation rejection such as kidney, heart, liver, lung, bone marrow, pancreas (islet cells), cornea, small bowel, and skin allografts, and heart valve xenografts; in the treatment or inhibition of autoimmune diseases such as lupus, rheumatoid arthritis, diabetes mellitus, myasthenia gravis, and multiple sclerosis; and diseases of inflammation such as psoriasis, dermatitis, eczema, seborrhea, inflammatory bowel disease, and eye uveitis.

The compounds, endogenous ligands and exogenous ligands mentioned above can also have antitumor, antifungal activities, and antiproliferative activities. The compounds of this invention therefore can be also useful in treating solid tumors, adult T-cell leukemia/lymphoma, fungal infections, and hyperproliferative vascular diseases such as restenosis and atherosclerosis.

EXAMPLE 1

The proteins of the present invention were isolated utilizing a fusion protein of glutathione S-transferase--FK506 binding protein12 (GST-FKBP). GST-FKBP is produced by a recombinant E. coli containing the plasmid, pGEX-FKBP. The cells were grown, induced with IPTG and the fusion protein was isolated using standard technology described in D.B. Smith and K.S. Johnson, Gene 67, 31, 1988 and K.L. Guan and J.E. Dixon, Anal. Biochem. 192, 262, 1991. The solution containing glutathione and GST-FKBP12 was exchanged 5x using a Centricon 10 filtration unit (Amicon) to remove the glutathione and exchange the buffer.

Molt 4 cells ($1x10^9$) were grown in standard media (RPMI 1640 containing 100 U/ml pennicillin, 100 ug/ml L-glutamine, 10% FCS). The cells were harvested and rinsed 3x with PBS (50mM phosphate buffer, pH 7.0, 150 mM NaCl), flash frozen in dry-ice ethanol bath and stored at -80°C. On ice, the cells were thawed and lysed using a dounce homogenizer with B pestle in 5 ml of buffer A (10 mM Hepes, pH 7.5, 20 mM KCl, 1 mM EDTA, 0.4 mM PMSF and 2 mM 2-ME). After the debris was cleared by centrifugation at 1600g for 10 min. and the membrane fraction was concentrated by 100,000g centrifugation (1 hour), the 100,000 g pellet was incubated in 3 ml buffer B (50 mM Tris, pH 7.2, 100 mM NaCl, 20 mM KCl, 0.2 mM PMSF, 1 mM 2-ME, 2 mM $CaCl_2$, 2 mM $MgCl_2$, 5 µg/ml aprotinin, leupeptin, pepstatin A and antipain), containing 12 mM CHAPSO for two hours at 4°C. The solubilized membrane proteins were separated from the debris by a 100,000 g centrifugation. After preabsorption of the supernatant for 3-18 hours with 0.4 ml glutathione sepharose resin swollen in buffer B, the supernatant was incubated with complexed Rapamycin-GST-FKBP12 (preformed by incubation of 660 ug GST-FKBP + 60 ug RAPA in buffer B for 1-2 hours, 4°C) for two hours at 4°C. The supernatant was then incubated for 2 hours at 4°C with 100 ul glutathione resin (1:1 Buffer B). Nonspecific proteins were rinsed 5x with buffer C (buffer B + 0.1% Triton x 100) and the proteins eluted from the resin in Laemli buffer by incubation at 95°C for 3 minutes and microcentrifugation. The proteins were separated by size using a 7% SDS-PAGE followed by silver stain. Four bands corresponding to proteins of molecular weights of 210kDa, 208 kDa, 148 kDa, and 125 kDa were present in higher concentrations in the sample containing RAPA + GST-FKBP12 vs GST-FKBP alone.

The mitogen-stimulated thymocyte proliferation assay called the LAF, can be inhibited by rapamycin or analogs such as demethoxyrapamycin and indicates relative activity of rapamycin analogs in immunosuppression. The same proteins were isolated using GST-FKBP complexed with the immunosuppressive analog, demethoxyrapamycin (Table1). The Diels Alder adducts bound to FKBP12 and inhibited PPIase activity of FKBP12 but did not exhibit detectable immunosuppressive activity and thus do not bind to the target of rapamycin. The use of these two compounds complexed with GST-FKBP12 in the analogous isolation procedure (ie. replacing rapamycin:GST-FKBP12) yielded background levels of the 210kDa proteins (no rapamycin)(Table 1). FK506, is an immunosuppressive compound which binds to

FKBP and and mediates at least some of its effects through the binding of the FK506-FKBP complex with calcineurin. FK506 when complexed with GST-FKBP in an analogous procedure yielded only background levels of the 210 kDa protein (Table 1).

TABLE 1

Comparison of Binding of Rapamycin Analog--FKBP12 complexes to 210 kDa Protein

Compound	210 kDa	LAF	PPIase(Ki)
RAPA	+++	6 nM	0.12nM
demethoxyrapamycin	+++	58nM	4.4 nM
Diels Alder adduct (phenyl)	±	>1000nM	12 nM
Diels Alder adduct (methyl)	±	>1000nM	12 nM
FK506	±	3nM*	0.4 nM
none (FKBP)	±		

(* mechanism of action is different)

It is known that rapamycin must bind to a member of the FKBP family in order to mediate its effects. To verify that the proteins of this invention bind to the complex RAPA-GST-FKBP and not individually to rapamycin or FKBP12, a modified isolation procedure was employed. The modification consists of using (1) a rapamycin-42-biotin glycinate ester in place of rapamycin (both exhibit equivalent immunosuppressive activity in the LAF assay), (2) no exogenous FKBP and (3) a strepatavidin-conjugated resin in place of glutathione-resin. Only background levels of the 210 kDa protein was isolated using this modified isolation procedure.

The 210 kDa protein was isolated using the GST--FKBP12--rapamycin complex from BJAB cells (B cell lymphoma) and normal human T lymphocytes purified by Ficoll-Hypaque and T cell columns.

The results of the partial amino acid composition analysis are set forth in Table 2, below. It should be noted that the percentage of the basic amino acids was not determined.

TABLE 2

Peak Number	Component Name	Retention Time	Peak Area	Response Factor	Peak Height	Concentration No./50ul
		9.38				
		11.09				
1	Asp/Asn	12.06	12.47076	0.02344	0.05142	0.30
2	Thr	13.05	2.92898	0.00000	0.00985	0.068
3	Ser	13.78	6.43968	0.00000	0.01995	0.15
		15.68				
4	Glu/Gln	16.87	25.47273	0.00000	0.05285	0.59
	Prp	18.24				0.14
5	Gly	22.35	21.50384	0.00000	0.04645	0.44
		22.90				
6	Ala	23.73	16.69160	0.00000	0.03113	0.36
		26.06				
		28.81				
7	Val	29.39	4.83196	0.00000	0.00605	0.11
	Met	32.28				
8	Ile	34.10	3.00560	0.2326	0.00782	0.0699
9	Leu	35.09	5.73202	0.02331	0.01372	0.1383
10	nLeu	36.27	20.48232	0.02174	0.04286	0.4453

TABLE 2 (Cont'd)

Peak Number	Component Name	Retention Time	Peak Area	Response Factor	Peak Height	Concentration No./50µl
11	Tyr	38.33	1.44792	0.02618	0.00226	0.0379
12	Phe	40.05	1.25017	0.02703	0.00187	0.0338
13	His	47.79	1.50905	0.02553	0.00580	0.0385
14		51.80	12.66136	0.00000	0.01960	0.0000
15	Lys	53.34	9.90767	0.02283	0.02274	0.2262
Totals			146.53645		0.33436	
Not Determined			144.29			

EXAMPLE 2

The 210 kDa (210±20 kDa) protein of this invention was isolated from 4 x 10^{11} Molt 4 cells using the affinity matrix protocol as described previously. Bound proteins were eluted from the affinity matrix with 1x Laemli buffer without glycerol and dye (0.0625 M Tris-HCl, ph6.8, 2% SDS, 0.37M b-mercaptoethanol) and were concentrated 3 consecutive times by centrifugation using centricon 100 (Amicon, Beverly, MA) at 4 °C the first two times and at 18 °C the third time. The concentrated sample was eluted from the centricon 100 filter by incubating 2 hours at room temperature with an equal volume of 2 x laemli buffer without glycerol and dye the first 2 x and 2 x laemli buffer the third time. The proteins in the sample were separated by PAGE on a 1.5mm thick 7% polyacrylamide gel (38:1). The proteins were transferred to polyvinylidine difluoride, PVDF, (Biorad, Hercules, CA) in 10 x Tris/glycine buffer (Biorad) containing 0.037% SDS at 50 mAmps at 4 °C overnight. The proteins on the PVDF were stained with amido black (Biorad) in 10% ethanol, 2% acetic acid and the appropriate band was excised, rinsed with PBS and water and stored frozen.

Sequencing

The protein (approx. 3 ug) on the PVDF membrane was digested in situ with trypsin using a modification described by J. Fernandez et al, (Anal.Biochem. 201: 255-64, 1992). Briefly, the PVDF was cut into 1 mm^2 pieces, prewet, and the protein digested in a 100mM Tris-HCl, pH buffer containing 10% acetonitrile, and 1% reduced triton (CalBiochem) with 0.2ug trypsin at 37 °C for 6 hours followed by addition of 0.2 ug trypsin and incubation overnight. The fragments were eluted from the membrane by sonication and the buffer containing the fragments were separated by microfuge centifugation. The membranes were backextracted 2x (i.e., 50 ul buffer was added to membranes, sonicated, and centrifuged in a microfuge and solution pooled with the original buffer containing the eluted fragments.) The sample (140-145 ul) was separated by narrow bore high performance liquid chromatography using a Vydac C18 2.1mm x 150 mm reverse phase column on a Hewlett Packard HPLC 1090 with a 40 diode array detector as described previously by W.Lane et al, (J.Protein Chem., 10(2): 151-60, 1991). Multiple fractions were collected and measured for absorption at multiple wavelengths (210, 277 and 292 nm). Optimal fractions were chosen for sequencing based on resolution, symmetry, and ultraviolet absorption and spectra (210 nm, 277 nm and 292 nm).. An aliquot (5%) of the optimal fractions was analyzed for homogeneity and length of fragment by matrix assisted laser desorption time of flight mass spectrometry, MALDE-TOF-MS, on a Finnigan lasermat. Selected optimal fractions were sequenced by automated Edman degradation on an Applied Biosystems 477A protein sequencer using microcartridge and manufacturer's recommended chemistry cycle.

Sequence comparison

Comparison was performed using the Intelligenetics suite (Intelligenetics, CA) .

Sequences

Utilizing the methods mentioned above, it was determined that the 210 kDa (210±20 kDa) protein of this invention contains peptide fragments, four of which have amino acid sequences as shown below:

a) ILLNIEHR;

B) LIRPYMEPILK;

c) DXMEAQE; and

d) QLDHPLPTVHPQVTYAYM(K)

Those skilled in the art will recognize the one-letter symbols for the amino acids in question (the definitions for which can also be seen at page 21 of the text *Biochemistry*, Third Edition, W.H. Freeman and Company, © 1988 by Lubert Stryer). Those so skilled will also understand that the X in sequence c) indicates an as yet unidentified amino acid and the parentheses in sequence d) indicates that the amino acid in the position in question is possibly lysine.

As mentioned previously, the present invention includes fragmented or truncated forms of the proteins mentioned herein. This includes proteins which have as part or all of their amino acid sequence one or more of the four sequences listed as a)-d), above. For the purposes of the claims, below, the proteins referred to as including one or more of the "internal amino acid sequences" are understood to be any protein which contains one of the sequences listed above, whether the protein is comprised wholly of one or more of the sequences a)-d) or whether one or more of the sequences mentioned above form any portion of the protein. This is understood to include all locations on the protein's amino acid sequence including, but not limited to, those sections of the protein which initiate and terminate the protein's amino acid chain.

These partial amino acid sequences were compared with sequences in the Genbank database. There was identity with the sequence, accession number L34075 (Brown et al., Nature 369, 756-758 (1994)). The cDNA of the SEP gene was cloned as follows: Two micrograms of Molt 4 cDNA (Clontech, Palo Alto,CA) in 1 x PCR buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1 mM MgCl, 200 µM dATP, 200 µM dTTP, 200 µM dCTP, 200 µM dGTP; Perkin Elmer,) with 1 unit Taq polymerase (Perkin Elmer),was amplified by Polymerase chain reaction (PCR) at 94 C for 30 sec., 66 C for 4 min for 30 cycles, 72 C for 10 min by three separate reactions containing one of the following pairs of oligomers:

CGATCGGTCGACTGCAGCACTTTGGGGATTGTGCTCTC and
GCGGCCGCAGCTTTCTTCATGCATGACAACAGCCCAGGC; or
GCGGCCGCAAGCTTCAAGTATGCAAGCCTGTGCGGCAAGA and
CGATCGGTCGACACCTTCTGCATCAGAGTCAAGTGGTCA; or
GCGGCCGCAAGCTTCCTCAGCTCACATCCTTAGAGCTGCA and
CGATCGGTCGACTTATTACCAGAAAGGGCACCAGCCAATATA .

The oligonucleotides were synthesized and isolated by methods previously described and known in the art (Chemical and Enzymatic Synthesis of Gene Fragments, ed. by H.G.Gassin and Anne Lang, Verlag Chemie, FLA, 1982). The resulting PCR products named SEP3, SEP4, and SEP5, respectively, were incubated at 15 C overnight in buffer containing T4 DNA ligase (1 unit) and 50 ng pcII which was modified to efficiently ligate PCR products (TA cloning kit, Invitrogen, San Diego, CA) to yield PCR-pcII ligated products. The PCR-pcII products were transformed into competent *E. coli* INValphaF cells obtained commercially from Invitrogen. Miniprep DNA was prepared using the Quiagen miniprep kits (Quiagen, Chatsworth, CA) and the clones containing the appropriate sized PCR product were identified by restriction enzyme digestion with commercially available HindIII or Sal I , electrophoresis, and comparison to standards. Sep2 and Sep1 cDNA was made using the TimeSaver cDNA synthesis Kit (Pharmacia, Piscataway, NJ) with the first strand synthesis reaction containing oligodT (0.13 μg) and 250 pmoles of

CGATCGGTCGACCAGATGAGCACATCATAGCGCTGATGA or
CGATCGGTCGACAAATTCAAAGCTGCCAAGCGTTCGGAG,

respectively. Sep2 and Sep1 second strand synthesis was performed using the TimeSaver cDNA synthesis kit with the addition of 250 pmoles of

GCGGCCGCAAGCTTTGGCTCGAGCAATGGGGCCAGGCA or
GCGGCCGCAAGCTTAAGATGCTTGGAACCGCACCTGCCG,

respectively. The Sep2 and Sep1 cDNA was then amplified by PCR using

CGATCGGTCGACCAGATGAGCACATCATAGCGCTGATGA and 19.
GCGGCCGCAAGCTTTGGCTCGAGCAATGGGGCCAGGCA or 20
GCGGCCGCAAGCTTAAGATGCTTGGAACCGCACCTGCCG and 21
CGATCGGTCGACAAATTCAAAGCTGCCAAGCGTTCGGAG, 22

respectively as described above. The Sep2 PCR products were cloned int the TA cloning kit (Invitrogen). The Sep 1 PCR products were digested with Hind III and Sal I, separated from the pcII vector by agarose electrophoresis. The Sep1 (HindIII-SalI) fragment was isolated using the Sephaglas bandprep kit from Pharmacia and cloned into the HindIII and Sal I sites of pUC19 as described (Sambrook et al., Molecular Cloning Cold Spring Harbor, 1989). Ligation of the isolated Sep2(HindIII, AspI) and Sep3(AspI, SalI) fragments or Sep4(HindIII, AccIII/MroI) and Sep5(AccIII/MroI, Sal I) fragments into pUC18(HindIII, SalI) vector and transformation of competent E. coli INValphaF cells (Invitrogen) was performed by techniques known to those skilled in the art (Sambrook et al., Molecular Cloning Cold Spring Harbor, 1989) to obtain pUC18-Sep 23 and pUC18-Sep45 which contain nucleotides 1468- 5326 and 4964 - 7653, respectively, of the full length clone shown in the attached Sequence No. 1. Ligation of the pUC19-Sep1 (EcoRV, SalI), Sep2345 (EcoRV, SalI) fragments and transformation of competent E. coli INValphaF cells (Invitrogen) were performed by techniques known to those skilled in the art (as described by Sambrook et al., Molecular Cloning Cold Spring Harbor, 1989) to obtain the full length clone. The nucleic acid sequence coding for this protein and its amino acid sequence are shown in Sequence No. 1.

A fusion protein, called glutathione S transferase-sirolimus effector protein, GST-SEP, was engineered by subcloning the Sep4 and Sep5 fragments into the plasmid, pGEX-KG (Guan, K. and Dixon, J.E. (1991) Anal. Biochem. 192, 262-267) as follows. Briefly, Sep4 was digested with commercially available HindIII restriction enzyme, the restriction site was filled in with the Klenow fragment of DNA polymerase (Gibco), and the DNA was extracted with phenol-chloroform and ethanol precipitated using techniques known by those skilled in the art (Sambrook et al., Molecular Cloning Cold Spring Harbor, 1989). The SEP4 (HindIII-Klenow) was further digested with MroI restriction enzyme, separated from the pcII vector by agarose electrophoresis and isolated as the fragment SEP4-HindIII-Klenow-MroI. Sep5 fragment was prepared by

digestion with SalI and MroI, separated from the pcII vector by agarose electrophoresis and isolated as the fragment SEP5-SalI-MroI. pGEX-KG (Guan, K. and Dixon, J.E. (1991) Anal. Biochem. 192, 262-267) was digested with Nco I, filled in with the Klenow fragment of DNA polymerase and the DNA was extracted with phenol-chloroform and ethanol precipitated, using techniques of those skilled in the art (Sambrook et al., Molecular Cloning Cold Spring Harbor, 1989). pGEX-KG (NcoI, Klenow) was further digested with Sal I, separated from the undigested vector by agarose electrophoresis and isolated as the vector pGEX-KG-NcoI-Klenow-SalI, using techniques of those skilled in the art. Ligation of the vector, pGEX-KG-NcoI-Klenow-SalI and Sep 4 (HindIII, MroI) and Sep5 (MroI, SalI) fragments and transformation into E. coli strain INValphaF cells (Invitrogen) using techniques of those skilled in the art yielded the plasmid, pGEX-Sep45. Other *E. coli hosts* such as BL21 can also be used The DNA and protein sequence of this fusion protein is shown in Sequence No. 2.

Flag sequences and kinase recognition domain of heart muscle kinase can be added at the amino terminal end, by methods known in the art (see Chen et al., *Gene* 1994 Feb. 11; 139 (1): 73-75) within SEP or at the carboxy terminus of SEP, SEP4,5 or other fragments using an oligonucleotide which includes the coding sequence for Asp Tyr Lys Asp Asp Asp Asp Lys. The fusion protein can be isolated by affinity chromatography with anti-flag specific antibodies using the commercially available kits from IBI, New Haven, Conn.

Transformed host cells containing sequences of this invention have been deposited with the American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, Maryland 20852, USA, and have been given the ATCC designations listed below:

Sequence	ATCC Designation
a) pUC19-Sep1(nucleotides 1- 1785 of Sequence No. 1)	ATCC 69756
b) pUC18-Sep23 (nucleotides 1468- 5326 of Sequence No. 1)	ATCC 69753

c) pUC18-Sep45 (nucleotides 4964 - 7653 of Sequence No. 1) ATCC 69754

d) pUC19-Sep1-5 (ATCC 69756 1-7653 of sequence 1) ATCC 69829

e) pGEX-Sep45 plasmids (Sequence 2) ATCC 69755.

EXAMPLE 3

The 210 kDa protein of this invention was also isolated by the techniques described in Example 1 utilizing the following rapamycin analogs:

a) 42-Deoxy-42-[1-(1,1-dimethylethoxy)-2-oxoethoxy] rapamycin (which is described in U.S. Pat. No. 5,233,036);

b) 42-[O-[(1,1-Dimethylethyl)dimethylsilyl]] rapamycin (described in U.S. Pat. No. 5,120,842);

c) Rapamycin 42-ester with N-[1,1-dimethylethoxy)carbonyl]-N-methylglycine (described in U.S. Pat. No. 5,130,307);

d) Rapamycin 42-ester with 5-(1,1-dimethylethoxy)-2-[[(1,1-dimethylethoxy)carbonyl]amino]-5-oxopentanoic acid ethyl acetate solvate three quarter hydrate (see U.S. Pat. No. 5,130,307);

e) Rapamycin 42-ester with N-[(1,1-dimethylethoxy)carbonyl]glycylglycine hydrate (see U.S. Pat. No. 5,130,307); and

f) Rapamycin 42-ester with N2, N6-bis[(1,1-dimethylethoxy)carbonyl]-L-lysine (see U.S. Pat. No. 5,130,307).

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Molnar-Kimber, Katherine L.
Failli, Amedeo F.
Caggiano, Thomas J.
Nakanishi, Koji
Chen, Yanqiu

(ii) TITLE OF INVENTION: Effector Proteins of Rapamycin

(iii) NUMBER OF SEQUENCES: 2

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Ronald W. Alice,
American Home Products Corporation
(B) STREET: 5 Giralda Farms
(C) CITY: Madison
(D) STATE: New Jersey
(E) COUNTRY: USA
(F) ZIP: 07940-0874

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage

(B) COMPUTER: Apple Macintosh

(C) OPERATING SYSTEM: Macintosh 7.1

(D) SOFTWARE: Microsoft Word

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/312,023
(B) FILING DATE: 26-SEPTEMBER-1994
(C) APPLICATION NO: US 08/207,975
(E) FILING DATE: 08-MARCH-1994

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Eck, Steven R.
(B) REGISTRATION NUMBER: 36,126
(C) REFERENCE/DOCKET NUMBER: AHP-93167-2-C2

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (610) 902-2628
(B) TELEFAX: (610) 688-0273

(2) INFORMATION FOR SEQ. ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7653
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double-stranded
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: sense orientation of double-stranded cDNA to mRNA

(iii) HYPOTHETICAL: no

(iv) ANTISENSE: no

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Molt 4 human T-cell leukemia cells
(B) STRAIN: ATCC Strain CRL 1582

(xi) SEQUENCE DESCRIPTION: SEQ. ID NO: 1

```
AAG ATG CTT GGA ACC GGA CCT GCC GCC GCC ACC ACC GCT GCC ACC ACA  48
    Met Leu Gly Thr Gly Pro Ala Ala Ala Thr Thr Ala Ala Thr Thr
      1               5                  10                  15

TCT AGC AAT GTG AGC GTC CTG CAG CAG TTT GCC AGT GGC CTA AAG AGC  96
Ser Ser Asn Val Ser Val Leu Gln Gln Phe Ala Ser Gly Leu Lys Ser
                  20                  25                  30

CGG AAT GAG GAA ACC AGG GCC AAA GCC GCC AAG GAG CTC CAG CAC TAT 144
Arg Asn Glu Glu Thr Arg Ala Lys Ala Ala Lys Glu Leu Gln His Tyr
              35                  40                  45

GTC ACC ATG GAA CTC CGA GAG ATG AGT CAA GAG GAG TCT ACT CGC TTC 192
Val Thr MET Glu Leu Arg Glu MET Ser Gln Glu Glu Ser Thr Arg Phe
          50                  55                  60

TAT GAC CAA CTG AAC CAT CAC ATT TTT GAA TTG GTT TCC AGC TCA GAT 240
Tyr Asp Gln Leu Asn His His Ile Phe Glu Leu Val Ser Ser Ser Asp
      65                  70                  75

GCC AAT GAG AGG AAA GGT GGC ATC TTG GCC ATA GCT AGC CTC ATA GGA 288
Ala Asn Glu Arg Lys Gly Gly Ile Leu Ala Ile Ala Ser Leu Ile Gly
  80                  85                  90                  95

GTG GAA GGT GGG AAT GCC ACC CGA ATT GGC AGA TTT GCC AAC TAT CTT 336
Val Glu Gly Gly Asn Ala Thr Arg Ile Gly Arg Phe Ala Asn Tyr Leu
                 100                 105                 110

CGG AAC CTC CTC CCC TCC AAT GAC CCA GTT GTC ATG GAA ATG GCA TCC 384
Arg Asn Leu Leu Pro Ser Asn Asp Pro Val Val MET Glu MET Ala Ser
             115                 120                 125
```

```
AAG GCC ATT GGC CGT CTT GCC ATG GCA GGG GAC ACT TTT ACC GCT GAG  432
Lys Ala Ile Gly Arg Leu Ala MET Ala Gly Asp Thr Phe Thr Ala Glu
        130                 135                 140

TAC GTG GAA TTT GAG GTG AAG CGA GCC CTG GAA TGG CTG GGT GCT GAC  480
Tyr Val Glu Phe Glu Val Lys Arg Ala Leu Glu Trp Leu Gly Ala Asp
    145                 150                 155

CGC AAT GAG GGC CGG AGA CAT GCA GCT GTC CTG GTT CTC CGT GAG CTG  528
Arg Asn Glu Gly Arg Arg His Ala Ala Val Leu Val Leu Arg Glu Leu
160                 165                 170                 175

GCC ATC AGC GTC CCT ACC TTC TTC TTC CAG CAA GTG CAA CCC TTC TTT  576
Ala Ile Ser Val Pro Thr Phe Phe Phe Gln Gln Val Gln Pro Phe Phe
                180                 185                 190

GAC AAC ATT TTT GTG GCC GTG TGG GAC CCC AAA CAG GCC ATC CGT GAG  624
Asp Asn Ile Phe Val Ala Val Trp Asp Pro Lys Gln Ala Ile Arg Glu
            195                 200                 205

GGA GCT GTA GCC GCC CTT CGT GCC TGT CTG ATT CTC ACA ACC CAG CGT  672
Gly Ala Val Ala Ala Leu Arg Ala Cys Leu Ile Leu Thr Thr Gln Arg
        210                 215                 220

GAG CCG AAG GAG ATG CAG AAG CCT CAG TGG TAC AGG CAC ACA TTT GAA  720
Glu Pro Lys Glu MET Gln Lys Pro Gln Trp Tyr Arg His Thr Phe Glu
    225                 230                 235

GAA GCA GAG AAG GGA TTT GAT GAG ACC TTG GCC AAA GAG AAG GGC ATG  768
Glu Ala Glu Lys Gly Phe Asp Glu Thr Leu Ala Lys Glu Lys Gly MET
240                 245                 250                 255

AAT CGG GAT GAT CGG ATC CAT GGA GCC TTG TTG ATC CTT AAC GAG CTG  816
Asn Arg Asp Asp Arg Ile His Gly Ala Leu Leu Ile Leu Asn Glu Leu
                260                 265                 270

GTC CGA ATC AGC AGC ATG GAG GGA GAG CGT CTG AGA GAA GAA ATG GAA  864
Val Arg Ile Ser Ser MET Glu Gly Glu Arg Leu Arg Glu Glu MET Glu
            275                 280                 285

GAA ATC ACA CAG CAG CAG CTG GTA CAC GAC AAG TAC TGC AAA GAT CTC  912
Glu Ile Thr Gln Gln Gln Leu Val His Asp Lys Tyr Cys Lys Asp Leu
        290                 295                 300

ATG GGC TTC GGA ACA AAA CCT CGT CAC ATT ACC CCC TTC ACC AGT TTC  960
MET Gly Phe Gly Thr Lys Pro Arg His Ile Thr Pro Phe Thr Ser Phe
    305                 310                 315

CAG GCT GTA CAG CCC CAG CAG TCA AAT GCC TTG GTG GGG CTG CTG GGG 1008
Gln Ala Val Gln Pro Gln Gln Ser Asn Ala Leu Val Gly Leu Leu Gly
320                 325                 330                 335
```

```
TAC AGC TCT CAC CAA GGC CTC ATG GGA TTT GGG ACC TCC CCC AGT CCA  1056
Tyr Ser Ser His Gln Gly Leu MET Gly Phe Gly Thr Ser Pro Ser Pro
            340                 345                 350

GCT AAG TCC ACC CTG GTG GAG AGC CGG TGT TGC AGA GAC TTG ATG GAG  1104
Ala Lys Ser Thr Leu Val Glu Ser Arg Cys Cys Arg Asp Leu MET Glu
        355                 360                 365

GAG AAA TTT GAT CAG GTG TGC CAG TGG GTG CTG AAA TGC AGG AAT AGC  1152
Glu Lys Phe Asp Gln Val Cys Gln Trp Val Leu Lys Cys Arg Asn Ser
    370                 375                 380

AAG AAC TCG CTG ATC CAA ATG ACA ATC CTT AAT TTG TTG CCC CGC TTG  1200
Lys Asn Ser Leu Ile Gln MET Thr Ile Leu Asn Leu Leu Pro Arg Leu
385                 390                 395

GCT GCA TTC CGA CCT TCT GCC TTC ACA GAT ACC CAG TAT CTC CAA GAT  1248
Ala Ala Phe Arg Pro Ser Ala Phe Thr Asp Thr Gln Tyr Leu Gln Asp
400                 405                 410                 415

ACC ATG AAC CAT GCC CTA AGC TGT GTC AAG AAG GAG AAG GAA CGT ACA  1296
Thr MET Asn His Ala Leu Ser Cys Val Lys Lys Glu Lys Glu Arg Thr
            420                 425                 430

GCG GCC TTC CAA GCC CTG GGG CTA CTT TCT GTG GCT GTG AGG TCT GAG  1344
Ala Ala Phe Gln Ala Leu Gly Leu Leu Ser Val Ala Val Arg Ser Glu
        435                 440                 445

TTT AAG GTC TAT TTG CCT CGC GTG CTG GAC ATC ATC CGA GCG GCC CTG  1392
Phe Lys Val Tyr Leu Pro Arg Val Leu Asp Ile Ile Arg Ala Ala Leu
    450                 455                 460

CCC CCA AAG GAC TTC GCC CAT AAG AGG CAG AAG GCA ATG CAG GTG GAC  1440
Pro Pro Lys Asp Phe Ala His Lys Arg Gln Lys Ala MET Gln Val Asp
465                 470                 475

GCC ACA GTC TTC ACT TGC ATC AGC ATG CTG GCT CGA GCA ATG GGG CCA  1488
Ala Thr Val Phe Thr Cys Ile Ser MET Leu Ala Arg Ala MET Gly Pro
480                 485                 490                 495

GGC ATC CAG CAG GAT ATC AAG GAG CTG CTG GAG CCC ATG CTG GCA GTG  1536
Gly Ile Gln Gln Asp Ile Lys Glu Leu Leu Glu Pro MET Leu Ala Val
            500                 505                 510

GGA CTA AGC CCT GCC CTC ACT GCA GTG CTC TAC GAC CTG AGC CGT CAG  1584
Gly Leu Ser Pro Ala Leu Thr Ala Val Leu Tyr Asp Leu Ser Arg Gln
        515                 520                 525

ATT CCA CAG CTA AAG AAG GAC ATT CAA GAT GGG CTA CTG AAA ATG CTG  1632
Ile Pro Gln Leu Lys Lys Asp Ile Gln Asp Gly Leu Leu Lys MET Leu
    530                 535                 540
```

```
TCC CTG GTC CTT ATG CAC AAA CCC CTT CGC CAC CCA GGC ATG CCC AAG  1680
Ser Leu Val Leu MET His Lys Pro Leu Arg His Pro Gly MET Pro Lys
    545                 550                 555

GGC CTG GCC CAT CAG CTG GCC TCT CCT GGC CTC ACG ACC CTC CCT GAG  1728
Gly Leu Ala His Gln Leu Ala Ser Pro Gly Leu Thr Thr Leu Pro Glu
560                 565                 570                 575

GCC AGC GAT GTG GGC AGC ATC ACT CTT GCC CTC CGA ACG CTT GGC AGC  1776
Ala Ser Asp Val Gly Ser Ile Thr Leu Ala Leu Arg Thr Leu Gly Ser
                580                 585                 590

TTT GAA TTT GAA GGC CAC TCT CTG ACC CAA TTT GTT CGC CAC TGT GCG  1824
Phe Glu Phe Glu Gly His Ser Leu Thr Gln Phe Val Arg His Cys Ala
            595                 600                 605

GAT CAT TTC CTG AAC AGT GAG CAC AAG GAG ATC CGC ATG GAG GCT GCC  1872
Asp His Phe Leu Asn Ser Glu His Lys Glu Ile Arg MET Glu Ala Ala
        610                 615                 620

CGC ACC TGC TCC CGC CTG CTC ACA CCC TCC ATC CAC CTC ATC AGT GGC  1920
Arg Thr Cys Ser Arg Leu Leu Thr Pro Ser Ile His Leu Ile Ser Gly
    625                 630                 635

CAT GCT CAT GTG GTT AGC CAG ACC GCA GTG CAA GTG GTG GCA GAT GTG  1968
His Ala His Val Val Ser Gln Thr Ala Val Gln Val Val Ala Asp Val
640                 645                 650                 655

CTT AGC AAA CTG CTC GTA GTT GGG ATA ACA GAT CCT GAC CCT GAC ATT  2016
Leu Ser Lys Leu Leu Val Val Gly Ile Thr Asp Pro Asp Pro Asp Ile
                660                 665                 670

CGC TAC TGT GTC TTG GCG TCC CTG GAC GAG CGC TTT GAT GCA CAC CTG  2064
Arg Tyr Cys Val Leu Ala Ser Leu Asp Glu Arg Phe Asp Ala His Leu
            675                 680                 685

GCC CAG GCG GAG AAC TTG CAG GCC TTG TTT GTG GCT CTG AAT GAC CAG  2112
Ala Gln Ala Glu Asn Leu Gln Ala Leu Phe Val Ala Leu Asn Asp Gln
        690                 695                 700

GTG TTT GAG ATC CGG GAG CTG GCC ATC TGC ACT GTG GGC CGA CTC AGT  2160
Val Phe Glu Ile Arg Glu Leu Ala Ile Cys Thr Val Gly Arg Leu Ser
    705                 710                 715

AGC ATG AAC CCT GCC TTT GTC ATG CCT TTC CTG CGC AAG ATG CTC ATC  2208
Ser MET Asn Pro Ala Phe Val MET Pro Phe Leu Arg Lys MET Leu Ile
720                 725                 730                 735

CAG ATT TTG ACA GAG TTG GAG CAC AGT GGG ATT GGA AGA ATC AAA GAG  2256
Gln Ile Leu Thr Glu Leu Glu His Ser Gly Ile Gly Arg Ile Lys Glu
                740                 745                 750
```

```
CAG AGT GCC CGC ATG CTG GGG CAC CTG GTC TCC AAT GCC CCC CGA CTC  2304
Gln Ser Ala Arg MET Leu Gly His Leu Val Ser Asn Ala Pro Arg Leu
            755                 760                 765

ATC CGC CCC TAC ATG GAG CCT ATT CTG AAG GCA TTA ATT TTG AAA CTG  2352
Ile Arg Pro Tyr MET Glu Pro Ile Leu Lys Ala Leu Ile Leu Lys Leu
        770                 775                 780

AAA GAT CCA GAC CCT GAT CCA AAC CCA GGT GTG ATC AAT AAT GTC CTG  2400
Lys Asp Pro Asp Pro Asp Pro Asn Pro Gly Val Ile Asn Asn Val Leu
785                 790                 795

GCA ACA ATA GGA GAA TTG GCA CAG GTT AGT GGC CTG GAA ATG AGG AAA  2448
Ala Thr Ile Gly Glu Leu Ala Gln Val Ser Gly Leu Glu MET Arg Lys
800                 805                 810                 815

TGG GTT GAT GAA CTT TTT ATT ATC ATC ATG GAC ATG CTC CAG GAT TCC  2496
Trp Val Asp Glu Leu Phe Ile Ile Ile MET Asp MET Leu Gln Asp Ser
            820                 825                 830

TCT TTG TTG GCC AAA AGG CAG GTG GCT CTG TGG ACC CTG GGA CAG TTG  2544
Ser Leu Leu Ala Lys Arg Gln Val Ala Leu Trp Thr Leu Gly Gln Leu
        835                 840                 845

GTG GCC AGC ACT GGC TAT GTA GTA GAG CCC TAC AGG AAG TAC CCT ACT  2592
Val Ala Ser Thr Gly Tyr Val Val Glu Pro Tyr Arg Lys Tyr Pro Thr
    850                 855                 860

TTG CTT GAG GTG CTA CTG AAT TTT CTG AAG ACT GAG CAG AAC CAG GGT  2640
Leu Leu Glu Val Leu Leu Asn Phe Leu Lys Thr Glu Gln Asn Gln Gly
865                 870                 875

ACA CGC AGA GAG GCC ATC CGT GTG TTA GGG CTT TTA GGG GCT TTG GAT  2688
Thr Arg Arg Glu Ala Ile Arg Val Leu Gly Leu Leu Gly Ala Leu Asp
880                 885                 890                 895

CCT TAC AAG CAC AAA GTG AAC ATT GGC ATG ATA GAC CAG TCC CGG GAT  2736
Pro Tyr Lys His Lys Val Asn Ile Gly MET Ile Asp Gln Ser Arg Asp
            900                 905                 910

GCC TCT GCT GTC AGC CTG TCA GAA TCC AAG TCA AGT CAG GAT TCC TCT  2784
Ala Ser Ala Val Ser Leu Ser Glu Ser Lys Ser Ser Gln Asp Ser Ser
        915                 920                 925

GAC TAT AGC ACT AGT GAA ATG CTG GTC AAC ATG GGA AAC TTG CCT CTG  2832
Asp Tyr Ser Thr Ser Glu MET Leu Val Asn MET Gly Asn Leu Pro Leu
    930                 935                 940

GAT GAG TTC TAC CCA GCT GTG TCC ATG GTG GCC CTG ATG CGG ATC TTC  2880
Asp Glu Phe Tyr Pro Ala Val Ser MET Val Ala Leu MET Arg Ile Phe
945                 950                 955
```

```
CGA GAC CAG TCA CTC TCT CAT CAT CAC ACC ATG GTT GTC CAG GCC ATC  2928
Arg Asp Gln Ser Leu Ser His His His Thr MET Val Val Gln Ala Ile
960                 965                 970                 975

ACC TTC ATC TTC AAG TCC CTG GGA CTC AAA TGT GTG CAG TTC CTG CCC  2976
Thr Phe Ile Phe Lys Ser Leu Gly Leu Lys Cys Val Gln Phe Leu Pro
                    980                 985                 990

CAG GTC ATG CCC ACG TTC CTT AAT GTC ATT CGA GTC TGT GAT GGG GCC  3024
Gln Val MET Pro Thr Phe Leu Asn Val Ile Arg Val Cys Asp Gly Ala
                995                1000                1005

ATC CGG GAA TTT TTG TTC CAG CAG CTG GGA ATG TTG GTG TCC TTT GTG  3072
Ile Arg Glu Phe Leu Phe Gln Gln Leu Gly MET Leu Val Ser Phe Val
           1010                1015                1020

AAG AGC CAC ATC AGA CCT TAT ATG GAT GAA ATA GTC ACC CTC ATG AGA  3120
Lys Ser His Ile Arg Pro Tyr MET Asp Glu Ile Val Thr Leu MET Arg
      1025                1030                1035

GAA TTC TGG GTC ATG AAC ACC TCA ATT CAG AGC ACG ATC ATT CTT CTC  3168
Glu Phe Trp Val MET Asn Thr Ser Ile Gln Ser Thr Ile Ile Leu Leu
1040                1045                1050                1055

ATT GAG CAA ATT GTG GTA GCT CTT GGG GGT GAA TTT AAG CTC TAC CTG  3216
Ile Glu Gln Ile Val Val Ala Leu Gly Gly Glu Phe Lys Leu Tyr Leu
                1060                1065                1070

CCC CAG CTG ATC CCA CAC ATG CTG CGT GTC TTC ATG CAT GAC AAC AGC  3264
Pro Gln Leu Ile Pro His MET Leu Arg Val Phe MET His Asp Asn Ser
           1075                1080                1085

CCA GGC CGC ATT GTC TCT ATC AAG TTA CTG GCT GCA ATC CAG CTG TTT  3312
Pro Gly Arg Ile Val Ser Ile Lys Leu Leu Ala Ala Ile Gln Leu Phe
      1090                1095                1100

GGC GCC AAC CTG GAT GAC TAC CTG CAT TTA CTG CTG CCT CCT ATT GTT  3360
Gly Ala Asn Leu Asp Asp Tyr Leu His Leu Leu Leu Pro Pro Ile Val
1105                1110                1115

AAG TTG TTT GAT GCC CCT GAA GCT CCA CTG CCA TCT CGA AAG GCA GCG  3408
Lys Leu Phe Asp Ala Pro Glu Ala Pro Leu Pro Ser Arg Lys Ala Ala
1120                1125                1130                1135

CTA GAG ACT GTG GAC CGC CTG ACG GAG TCC CTG GAT TTC ACT GAC TAT  3456
Leu Glu Thr Val Asp Arg Leu Thr Glu Ser Leu Asp Phe Thr Asp Tyr
                1140                1145                1150

GCC TCC CGG ATC ATT CAC CCT ATT GTT CGA ACA CTG GAC CAG AGC CCA  3504
Ala Ser Arg Ile Ile His Pro Ile Val Arg Thr Leu Asp Gln Ser Pro
           1155                1160                1165
```

```
GAA CTG CGC TCC ACA GCC ATG GAC ACG CTG TCT TCA CTT GTT TTT CAG  3552
Glu Leu Arg Ser Thr Ala MET Asp Thr Leu Ser Ser Leu Val Phe Gln
        1170                1175                1180

CTG GGG AAG AAG TAC CAA ATT TTC ATT CCA ATG GTG AAT AAA GTT CTG  3600
Leu Gly Lys Lys Tyr Gln Ile Phe Ile Pro MET Val Asn Lys Val Leu
    1185                1190                1195

GTG CGA CAC CGA ATC AAT CAT CAG CGC TAT GAT GTG CTC ATC TGC AGA  3648
Val Arg His Arg Ile Asn His Gln Arg Tyr Asp Val Leu Ile Cys Arg
1200                1205                1210                1215

ATT GTC AAG GGA TAC ACA CTT GCT GAT GAA GAG GAG GAT CCT TTG ATT  3696
Ile Val Lys Gly Tyr Thr Leu Ala Asp Glu Glu Glu Asp Pro Leu Ile
                1220                1225                1230

TAC CAG CAT CGG ATG CTT AGG AGT GGC CAA GGG GAT GCA TTG GCT AGT  3744
Tyr Gln His Arg MET Leu Arg Ser Gly Gln Gly Asp Ala Leu Ala Ser
            1235                1240                1245

GGA CCA GTG GAA ACA GGA CCC ATG AAG AAA CTG CAC GTC AGC ACC ATC  3792
Gly Pro Val Glu Thr Gly Pro MET Lys Lys Leu His Val Ser Thr Ile
        1250                1255                1260

AAC CTC CAA AAG GCC TGG GGC GCT GCC AGG AGG GTC TCC AAA GAT GAC  3840
Asn Leu Gln Lys Ala Trp Gly Ala Ala Arg Arg Val Ser Lys Asp Asp
    1265                1270                1275

TGG CTG GAA TGG CTG AGA CGG CTG AGC CTG GAG CTG CTG AAG GAC TCA  3888
Trp Leu Glu Trp Leu Arg Arg Leu Ser Leu Glu Leu Leu Lys Asp Ser
1280                1285                1290                1295

TCA TCG CCC TCC CTG CGC TCC TGC TGG GCC CTG GCA CAG GCC TAC AAC  3936
Ser Ser Pro Ser Leu Arg Ser Cys Trp Ala Leu Ala Gln Ala Tyr Asn
                1300                1305                1310

CCG ATG GCC AGG GAT CTC TTC AAT GCT GCA TTT GTG TCC TGC TGG TCT  3984
Pro MET Ala Arg Asp Leu Phe Asn Ala Ala Phe Val Ser Cys Trp Ser
            1315                1320                1325

GAA CTG AAT GAA GAT CAA CAG GAT GAG CTC ATC AGA AGC ATC GAG TTG  4032
Glu Leu Asn Glu Asp Gln Gln Asp Glu Leu Ile Arg Ser Ile Glu Leu
        1330                1335                1340

GCC CTC ACC TCA CAA GAC ATC GCT GAA GTC ACA CAG ACC CTC TTA AAC  4080
Ala Leu Thr Ser Gln Asp Ile Ala Glu Val Thr Gln Thr Leu Leu Asn
    1345                1350                1355

TTG GCT GAA TTC ATG GAA CAC AGT GAC AAG GGC CCC CTG CCA CTG AGA  4128
Leu Ala Glu Phe MET Glu His Ser Asp Lys Gly Pro Leu Pro Leu Arg
1360                1365                1370                1375
```

```
GAT GAC AAT GGC ATT GTT CTG CTG GGT GAG AGA GCT GCC AAG TGC CGA  4176
Asp Asp Asn Gly Ile Val Leu Leu Gly Glu Arg Ala Ala Lys Cys Arg
                1380                1385                1390

GCA TAT GCC AAA GCA CTA CAC TAC AAA GAA CTG GAG TTC CAG AAA GGC  4224
Ala Tyr Ala Lys Ala Leu His Tyr Lys Glu Leu Glu Phe Gln Lys Gly
            1395                1400                1405

CCC ACC CCT GCC ATT CTA GAA TCT CTC ATC AGC ATT AAT AAT AAG CTA  4272
Pro Thr Pro Ala Ile Leu Glu Ser Leu Ile Ser Ile Asn Asn Lys Leu
        1410                1415                1420

CAG CAG CCG GAG GCA GCG GCC GGA GTG TTA GAA TAT GCC ATG AAA CAC  4320
Gln Gln Pro Glu Ala Ala Ala Gly Val Leu Glu Tyr Ala MET Lys His
    1425                1430                1435

TTT GGA GAG CTG GAG ATC CAG GCT ACC TGG TAT GAG AAA CTG CAC GAG  4368
Phe Gly Glu Leu Glu Ile Gln Ala Thr Trp Tyr Glu Lys Leu His Glu
1440                1445                1450                1455

TGG GAG GAT GCC CTT GTG GCC TAT GAC AAG AAA ATG GAC ACC AAC AAG  4416
Trp Glu Asp Ala Leu Val Ala Tyr Asp Lys Lys MET Asp Thr Asn Lys
                1460                1465                1470

GAC GAC CCA GAG CTG ATG CTG GGC CGC ATG CGC TGC CTC GAG GCC TTG  4464
Asp Asp Pro Glu Leu MET Leu Gly Arg MET Arg Cys Leu Glu Ala Leu
            1475                1480                1485

GGG GAA TGG GGT CAA CTC CAC CAG CAG TGC TGT GAA AAG TGG ACC CTG  4512
Gly Glu Trp Gly Gln Leu His Gln Gln Cys Cys Glu Lys Trp Thr Leu
        1490                1495                1500

GTT AAT GAT GAG ACC CAA GCC AAG ATG GCC CGG ATG GCT GCT GCA GCT  4560
Val Asn Asp Glu Thr Gln Ala Lys MET Ala Arg MET Ala Ala Ala Ala
    1505                1510                1515

GCA TGG GGT TTA GGT CAG TGG GAC AGC ATG GAA GAA TAC ACC TGT ATG  4608
Ala Trp Gly Leu Gly Gln Trp Asp Ser MET Glu Glu Tyr Thr Cys MET
1520                1525                1530                1535

ATC CCT CGG GAC ACC CAT GAT GGG GCA TTT TAT AGA GCT GTG CTG GCA  4656
Ile Pro Arg Asp Thr His Asp Gly Ala Phe Tyr Arg Ala Val Leu Ala
                1540                1545                1550

CTG CAT CAG GAC CTC TTC TCC TTG GCA CAA CAG TGC ATT GAC AAG GCC  4704
Leu His Gln Asp Leu Phe Ser Leu Ala Gln Gln Cys Ile Asp Lys Ala
            1555                1560                1565

AGG GAC CTG CTG GAT GCT GAA TTA ACT GCA ATG GCA GGA GAG AGT TAC  4752
Arg Asp Leu Leu Asp Ala Glu Leu Thr Ala MET Ala Gly Glu Ser Tyr
        1570                1575                1580
```

```
AGT CGG GCA TAT GGG GCC ATG GTT TCT TGC CAC ATG CTG TCC GAG CTG  4800
Ser Arg Ala Tyr Gly Ala MET Val Ser Cys His MET Leu Ser Glu Leu
    1585                1590                1595

GAG GAG GTT ATC CAG TAC AAA CTT GTC CCC GAG CGA CGA GAG ATC ATC  4848
Glu Glu Val Ile Gln Tyr Lys Leu Val Pro Glu Arg Arg Glu Ile Ile
1600                1605                1610                1615

CGC CAG ATC TGG TGG GAG AGA CTG CAG GGC TGC CAG CGT ATC GTA GAG  4896
Arg Gln Ile Trp Trp Glu Arg Leu Gln Gly Cys Gln Arg Ile Val Glu
                1620                1625                1630

GAC TGG CAG AAA ATC CTT ATG GTG CGG TCC CTT GTG GTC AGC CCT CAT  4944
Asp Trp Gln Lys Ile Leu MET Val Arg Ser Leu Val Val Ser Pro His
            1635                1640                1645

GAA GAC ATG AGA ACC TGG CTC AAG TAT GCA AGC CTG TGC GGC AAG AGT  4992
Glu Asp MET Arg Thr Trp Leu Lys Tyr Ala Ser Leu Cys Gly Lys Ser
        1650                1655                1660

GGC AGG CTG GCT CTT GCT CAT AAA ACT TTA GTG TTG CTC CTG GGA GTT  5040
Gly Arg Leu Ala Leu Ala His Lys Thr Leu Val Leu Leu Leu Gly Val
    1665                1670                1675

GAT CCG TCT CGG CAA CTT GAC CAT CCT CTG CCA ACA GTT CAC CCT CAG  5088
Asp Pro Ser Arg Gln Leu Asp His Pro Leu Pro Thr Val His Pro Gln
1680                1685                1690                1695

GTG ACC TAT GCC TAC ATG AAA AAC ATG TGG AAG AGT GCC CGC AAG ATC  5136
Val Thr Tyr Ala Tyr MET Lys Asn MET Trp Lys Ser Ala Arg Lys Ile
                1700                1705                1710

GAT GCC TTC CAG CAC ATG CAG CAT TTT GTC CAG ACC ATG CAG CAA CAG  5184
Asp Ala Phe Gln His MET Gln His Phe Val Gln Thr MET Gln Gln Gln
            1715                1720                1725

GCC CAG CAT GCC ATC GCT ACT GAG GAC CAG CAG CAT AAG CAG GAA CTG  5232
Ala Gln His Ala Ile Ala Thr Glu Asp Gln Gln His Lys Gln Glu Leu
        1730                1735                1740

CAC AAG CTC ATG GCC CGA TGC TTC CTG AAA CTT GGA GAG TGG CAG CTG  5280
His Lys Leu MET Ala Arg Cys Phe Leu Lys Leu Gly Glu Trp Gln Leu
    1745                1750                1755

AAT CTA CAG GGC ATC AAT GAG AGC ACA ATC CCC AAA GTG CTG CAG TAC  5328
Asn Leu Gln Gly Ile Asn Glu Ser Thr Ile Pro Lys Val Leu Gln Tyr
1760                1765                1770                1775

TAC AGC GCC GCC ACA GAG CAC GAC CGC AGC TGG TAC AAG GCC TGG CAT  5376
Tyr Ser Ala Ala Thr Glu His Asp Arg Ser Trp Tyr Lys Ala Trp His
                1780                1785                1790
```

```
GCG TGG GCA GTG ATG AAC TTC GAA GCT GTG CTA CAC TAC AAA CAT CAG  5424
Ala Trp Ala Val MET Asn Phe Glu Ala Val Leu His Tyr Lys His Gln
            1795                1800                1805

AAC CAA GCC CGC GAT GAG AAG AAG AAA CTG CGT CAT GCC AGC GGG GCC  5472
Asn Gln Ala Arg Asp Glu Lys Lys Lys Leu Arg His Ala Ser Gly Ala
        1810                1815                1820

AAC ATC ACC AAC GCC ACC ACT GCC GCC ACC ACG GCC GCC ACT GCC ACC  5520
Asn Ile Thr Asn Ala Thr Thr Ala Ala Thr Thr Ala Ala Thr Ala Thr
    1825                1830                1835

ACC ACT GCC AGC ACC GAG GGC AGC AAC AGT GAG AGC GAG GCC GAG AGC  5568
Thr Thr Ala Ser Thr Glu Gly Ser Asn Ser Glu Ser Glu Ala Glu Ser
1840                1845                1850                1855

ACC GAG AAC AGC CCC ACC CCA TCG CCG CTG CAG AAG AAG GTC ACT GAG  5616
Thr Glu Asn Ser Pro Thr Pro Ser Pro Leu Gln Lys Lys Val Thr Glu
            1860                1865                1870

GAT CTG TCC AAA ACC CTC CTG ATG TAC ACG GTG CCT GCC GTC CAG GGC  5664
Asp Leu Ser Lys Thr Leu Leu MET Tyr Thr Val Pro Ala Val Gln Gly
            1875                1880                1885

TTC TTC CGT TCC ATC TCC TTG TCA CGA GGC AAC AAC CTC CAG GAT ACA  5712
Phe Phe Arg Ser Ile Ser Leu Ser Arg Gly Asn Asn Leu Gln Asp Thr
        1890                1895                1900

CTC AGA GTT CTC ACC TTA TGG TTT GAT TAT GGT CAC TGG CCA GAT GTC  5760
Leu Arg Val Leu Thr Leu Trp Phe Asp Tyr Gly His Trp Pro Asp Val
    1905                1910                1915

AAT GAG GCC TTA GTG GAG GGG GTG AAA GCC ATC CAG ATT GAT ACC TGG  5808
Asn Glu Ala Leu Val Glu Gly Val Lys Ala Ile Gln Ile Asp Thr Trp
1920                1925                1930                1935

CTA CAG GTT ATA CCT CAG CTC ATT GCA AGA ATT GAT ACG CCC AGA CCC  5856
Leu Gln Val Ile Pro Gln Leu Ile Ala Arg Ile Asp Thr Pro Arg Pro
            1940                1945                1950

TTG GTG GGA CGT CTC ATT CAC CAG CTT CTC ACA GAC ATT GGT CGG TAC  5904
Leu Val Gly Arg Leu Ile His Gln Leu Leu Thr Asp Ile Gly Arg Tyr
            1955                1960                1965

CAC CCC CAG GCC CTC ATC TAC CCA CTG ACA GTG GCT TCT AAG TCT ACC  5952
His Pro Gln Ala Leu Ile Tyr Pro Leu Thr Val Ala Ser Lys Ser Thr
        1970                1975                1980

ACG ACA GCC CGG CAC AAT GCA GCC AAC AAG ATT CTG AAG AAC ATG TGT  6000
Thr Thr Ala Arg His Asn Ala Ala Asn Lys Ile Leu Lys Asn MET Cys
    1985                1990                1995
```

```
GAG CAC AGC AAC ACC CTG GTC CAG CAG GCC ATG ATG GTG AGC GAG GAG  6048
Glu His Ser Asn Thr Leu Val Gln Gln Ala MET MET Val Ser Glu Glu
2000                2005                2010                2015

CTG ATC CGA GTG GCC ATC CTC TGG CAT GAG ATG TGG CAT GAA GGC CTG  6096
Leu Ile Arg Val Ala Ile Leu Trp His Glu MET Trp His Glu Gly Leu
                    2020                2025                2030

GAA GAG GCA TCT CGT TTG TAC TTT GGG GAA AGG AAC GTG AAA GGC ATG  6144
Glu Glu Ala Ser Arg Leu Tyr Phe Gly Glu Arg Asn Val Lys Gly MET
                2035                2040                2045

TTT GAG GTG CTG GAG CCC TTG CAT GCT ATG ATG GAA CGG GGC CCC CAG  6192
Phe Glu Val Leu Glu Pro Leu His Ala MET MET Glu Arg Gly Pro Gln
        2050                2055                2060

ACT CTG AAG GAA ACA TCC TTT AAT CAG GCC TAT GGT CGA GAT TTA ATG  6240
Thr Leu Lys Glu Thr Ser Phe Asn Gln Ala Tyr Gly Arg Asp Leu MET
    2065                2070                2075

GAG GCC CAA GAG TGG TGC AGG AAG TAC ATG AAA TCA GGG AAT GTC AAG  6288
Glu Ala Gln Glu Trp Cys Arg Lys Tyr MET Lys Ser Gly Asn Val Lys
2080                2085                2090                2095

GAC CTC ACC CAA GCC TGG GAC CTC TAT TAT CAT GTG TTC CGA CGA ATC  6336
Asp Leu Thr Gln Ala Trp Asp Leu Tyr Tyr His Val Phe Arg Arg Ile
                2100                2105                2110

TCA AAG CAG CTG CCT CAG CTC ACA TCC TTA GAG CTG CAA TAT GTT TCC  6384
Ser Lys Gln Leu Pro Gln Leu Thr Ser Leu Glu Leu Gln Tyr Val Ser
            2115                2120                2125

CCA AAA CTT CTG ATG TGC CGG GAC CTT GAA TTG GCT GTG CCA GGA ACA  6432
Pro Lys Leu Leu MET Cys Arg Asp Leu Glu Leu Ala Val Pro Gly Thr
        2130                2135                2140

TAT GAC CCC AAC CAG CCA ATC ATT CGC ATT CAG TCC ATA GCA CCG TCT  6480
Tyr Asp Pro Asn Gln Pro Ile Ile Arg Ile Gln Ser Ile Ala Pro Ser
    2145                2150                2155

TTG CAA GTC ATC ACA TCC AAG CAG AGG CCC CGG AAA TTG ACA CTT ATG  6528
Leu Gln Val Ile Thr Ser Lys Gln Arg Pro Arg Lys Leu Thr Leu MET
2160                2165                2170                2175

GGC AGC AAC GGA CAT GAG TTT GTT TTC CTT CTA AAA GGC CAT GAA GAT  6576
Gly Ser Asn Gly His Glu Phe Val Phe Leu Leu Lys Gly His Glu Asp
                2180                2185                2190

CTG CGC CAG GAT GAG CGT GTG ATG CAG CTC TTC GGC CTG GTT AAC ACC  6624
Leu Arg Gln Asp Glu Arg Val MET Gln Leu Phe Gly Leu Val Asn Thr
            2195                2200                2205
```

```
CTT CTG GCC AAT GAC CCA ACA TCT CTT CGG AAA AAC CTC AGC ATC CAG  6672
Leu Leu Ala Asn Asp Pro Thr Ser Leu Arg Lys Asn Leu Ser Ile Gln
        2210                2215                2220

AGA TAC GCT GTC ATC CCT TTA TCG ACC AAC TCG GGC CTC ATT GGC TGG  6720
Arg Tyr Ala Val Ile Pro Leu Ser Thr Asn Ser Gly Leu Ile Gly Trp
    2225                2230                2235

GTT CCC CAC TGT GAC ACA CTG CAC GCC CTC ATC CGG GAC TAC AGG GAG  6768
Val Pro His Cys Asp Thr Leu His Ala Leu Ile Arg Asp Tyr Arg Glu
2240                2245                2250                2255

AAG AAG AAG ATC CTT CTC AAC ATC GAG CAT CGC ATC ATG TTG CGG ATG  6816
Lys Lys Lys Ile Leu Leu Asn Ile Glu His Arg Ile MET Leu Arg MET
                2260                2265                2270

GCT CCG GAC TAT GAC CAC TTG ACT CTG ATG CAG AAG GTG GAG GTG TTT  6864
Ala Pro Asp Tyr Asp His Leu Thr Leu MET Gln Lys Val Glu Val Phe
            2275                2280                2285

GAG CAT GCC GTC AAT AAT ACA GCT GGG GAC GAC CTG GCC AAG CTG CTG  6912
Glu His Ala Val Asn Asn Thr Ala Gly Asp Asp Leu Ala Lys Leu Leu
        2290                2295                2300

TGG CTG AAA AGC CCC AGC TCC GAG GTG TGG TTT GAC CGA AGA ACC AAT  6960
Trp Leu Lys Ser Pro Ser Ser Glu Val Trp Phe Asp Arg Arg Thr Asn
    2305                2310                2315

TAT ACC CGT TCT TTA GCG GTC ATG TCA ATG GTT GGG TAT ATT TTA GGC  7008
Tyr Thr Arg Ser Leu Ala Val MET Ser MET Val Gly Tyr Ile Leu Gly
2320                2325                2330                2335

CTG GGA GAT AGA CAC CCA TCC AAC CTG ATG CTG GAC CGT CTG AGT GGG  7056
Leu Gly Asp Arg His Pro Ser Asn Leu MET Leu Asp Arg Leu Ser Gly
                2340                2345                2350

AAG ATC CTG CAC ATT GAC TTT GGG GAC TGC TTT GAG GTT GCT ATG ACC  7104
Lys Ile Leu His Ile Asp Phe Gly Asp Cys Phe Glu Val Ala MET Thr
            2355                2360                2365

CGA GAG AAG TTT CCA GAG AAG ATT CCA TTT AGA CTA ACA AGA ATG TTG  7152
Arg Glu Lys Phe Pro Glu Lys Ile Pro Phe Arg Leu Thr Arg MET Leu
        2370                2375                2380

ACC AAT GCT ATG GAG GTT ACA GGC CTG GAT GGC AAC TAC AGA ATC ACA  7200
Thr Asn Ala MET Glu Val Thr Gly Leu Asp Gly Asn Tyr Arg Ile Thr
    2385                2390                2395

TGC CAC ACA GTG ATG GAG GTG CTG CGA GAG CAC AAG GAC AGT GTC ATG  7248
Cys His Thr Val MET Glu Val Leu Arg Glu His Lys Asp Ser Val MET
2400                2405                2410                2415
```

```
GCC GTG CTG GAA GCC TTT GTC TAT GAC CCC TTG CTG AAC TGG AGG CTG  7296
Ala Val Leu Glu Ala Phe Val Tyr Asp Pro Leu Leu Asn Trp Arg Leu
                2420                2425                2430

ATG GAC ACA AAT ACC AAA GGC AAC AAG CGA TCC CGA ACG AGG ACG GAT  7344
MET Asp Thr Asn Thr Lys Gly Asn Lys Arg Ser Arg Thr Arg Thr Asp
            2435                2440                2445

TCC TAC TCT GCT GGC CAG TCA GTC GAA ATT TTG GAC GGT GTG GAA CTT  7392
Ser Tyr Ser Ala Gly Gln Ser Val Glu Ile Leu Asp Gly Val Glu Leu
        2450                2455                2460

GGA GAG CCA GCC CAT AAG AAA ACG GGG ACC ACA GTG CCA GAA TCT ATT  7440
Gly Glu Pro Ala His Lys Lys Thr Gly Thr Thr Val Pro Glu Ser Ile
    2465                2470                2475

CAT TCT TTC ATT GGA GAC GGT TTG GTG AAA CCA GAG GCC CTA AAT AAG  7488
His Ser Phe Ile Gly Asp Gly Leu Val Lys Pro Glu Ala Leu Asn Lys
2480                2485                2490                2495

AAA GCT ATC CAG ATT ATT AAC AGG GTT CGA GAT AAG CTC ACT GGT CGG  7536
Lys Ala Ile Gln Ile Ile Asn Arg Val Arg Asp Lys Leu Thr Gly Arg
                2500                2505                2510

GAC TTC TCT CAT GAT GAC ACT TTG GAT GTT CCA ACG CAA GTT GAG CTG  7584
Asp Phe Ser His Asp Asp Thr Leu Asp Val Pro Thr Gln Val Glu Leu
            2515                2520                2525

CTC ATC AAA CAA GCG ACA TCC CAT GAA AAC CTC TGC CAG TGC TAT ATT  7632
Leu Ile Lys Gln Ala Thr Ser His Glu Asn Leu Cys Gln Cys Tyr Ile
        2530                2535                2540

GGC TGG TAC CCT TTC TGG TAA                                     7653
Gly Trp Tyr Pro Phe Trp
    2545
```

(3) INFORMATION FOR SEQ. ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3423
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double-stranded
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: sense orientation of double-stranded cDNA to mRNA
- (A) DESCRIPTION: Sequence No. 2 illustrates a GST-SEP45 fusion protein beginning at the first amino acid of the GST-SEP45 protein.

(iii) HYPOTHETICAL: no

(iv) ANTISENSE: no

(vi) ORIGINAL SOURCE:
- (A) ORGANISM: Molt 4 human T-cell leukemia cells
- (B) STRAIN: ATCC Strain CRL 1582

(xi) SEQUENCE DESCRIPTION: SEQ. ID NO: 2

```
ATG TCC CCT ATA CTA GGT TAT TGG AAA ATT AAG GGC CTT GTG CAA CCC  48
MET Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro
 1               5                  10                  15

ACT CGA CTT CTT TTG GAA TAT CTT GAA GAA AAA TAT GAA GAG CAT TTG  96
Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu
            20                  25                  30

TAT GAG CGC GAT GAA GGT GAT AAA TGG CGA AAC AAA AAG TTT GAA TTG 144
Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu
        35                  40                  45

GGT TTG GAG TTT CCC AAT CTT CCT TAT TAT ATT GAT GGT GAT GTT AAA 192
Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys
    50                  55                  60

TTA ACA CAG TCT ATG GCC ATC ATA CGT TAT ATA GCT GAC AAG CAC AAC 240
Leu Thr Gln Ser MET Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn
65                  70                  75                  80

ATG TTG GGT GGT TGT CCA AAA GAG CGT GCA GAG ATT TCA ATG CTT GAA 288
MET Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser MET Leu Glu
                85                  90                  95

GGA GCG GTT TTG GAT ATT AGA TAC GGT GTT TCG AGA ATT GCA TAT AGT 336
Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser
            100                 105                 110

AAA GAC TTT GAA ACT CTC AAA GTT GAT TTT CTT AGC AAG CTA CCT GAA 384
Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu
        115                 120                 125

ATG CTG AAA ATG TTC GAA GAT CGT TTA TGT CAT AAA ACA TAT TTA AAT 432
MET Leu Lys MET Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn
    130                 135                 140

GGT GAT CAT GTA ACC CAT CCT GAC TTC ATG TTG TAT GAC GCT CTT GAT 480
Gly Asp His Val Thr His Pro Asp Phe MET Leu Tyr Asp Ala Leu Asp
145                 150                 155                 160

GTT GTT TTA TAC ATG GAC CCA ATG TGC CTG GAT GCG TTC CCA AAA TTA 528
Val Val Leu Tyr MET Asp Pro MET Cys Leu Asp Ala Phe Pro Lys Leu
                165                 170                 175

GTT TGT TTT AAA AAA CGT ATT GAA GCT ATC CCA CAA ATT GAT AAG TAC 576
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr
            180                 185                 190

TTG AAA TCC AGC AAG TAT ATA GCA TGG CCT TTG CAG GGC TGG CAA GCC 624
Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala
        195                 200                 205
```

```
ACG TTT GGT GGT GGC GAC CAT CCT CCA AAA TCG GAT CTG GTT CCG CGT  672
Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg
    210                 215                 220

GGT GGA TCC CCG GGA ATT TCC GGT GGT GGT GGT GGT GGA ATT CTA GAC  720
Gly Gly Ser Pro Gly Ile Ser Gly Gly Gly Gly Gly Gly Ile Leu Asp
225                 230                 235                 240

GAC TCC ATG AGC TTC AAG TAT GCA AGC CTG TGC GGC AAG AGT GGC AGG  768
Asp Ser MET Ser Phe Lys Tyr Ala Ser Leu Cys Gly Lys Ser Gly Arg
                245                 250                 255

CTG GCT CTT GCT CAT AAA ACT TTA GTG TTG CTC CTG GGA GTT GAT CCG  816
Leu Ala Leu Ala His Lys Thr Leu Val Leu Leu Leu Gly Val Asp Pro
            260                 265                 270

TCT CGG CAA CTT GAC CAT CCT CTG CCA ACA GTT CAC CCT CAG GTG ACC  864
Ser Arg Gln Leu Asp His Pro Leu Pro Thr Val His Pro Gln Val Thr
        275                 280                 285

TAT GCC TAC ATG AAA AAC ATG TGG AAG AGT GCC CGC AAG ATC GAT GCC  912
Tyr Ala Tyr MET Lys Asn MET Trp Lys Ser Ala Arg Lys Ile Asp Ala
    290                 295                 300

TTC CAG CAC ATG CAG CAT TTT GTC CAG ACC ATG CAG CAA CAG GCC CAG  960
Phe Gln His MET Gln His Phe Val Gln Thr MET Gln Gln Gln Ala Gln
305                 310                 315                 320

CAT GCC ATC GCT ACT GAG GAC CAG CAG CAT AAG CAG GAA CTG CAC AAG 1008
His Ala Ile Ala Thr Glu Asp Gln Gln His Lys Gln Glu Leu His Lys
                325                 330                 335

CTC ATG GCC CGA TGC TTC CTG AAA CTT GGA GAG TGG CAG CTG AAT CTA 1056
Leu MET Ala Arg Cys Phe Leu Lys Leu Gly Glu Trp Gln Leu Asn Leu
            340                 345                 350

CAG GGC ATC AAT GAG AGC ACA ATC CCC AAA GTG CTG CAG TAC TAC AGC 1104
Gln Gly Ile Asn Glu Ser Thr Ile Pro Lys Val Leu Gln Tyr Tyr Ser
        355                 360                 365

GCC GCC ACA GAG CAC GAC CGC AGC TGG TAC AAG GCC TGG CAT GCG TGG 1152
Ala Ala Thr Glu His Asp Arg Ser Trp Tyr Lys Ala Trp His Ala Trp
    370                 375                 380

GCA GTG ATG AAC TTC GAA GCT GTG CTA CAC TAC AAA CAT CAG AAC CAA 1200
Ala Val MET Asn Phe Glu Ala Val Leu His Tyr Lys His Gln Asn Gln
385                 390                 395                 400

GCC CGC GAT GAG AAG AAG AAA CTG CGT CAT GCC AGC GGG GCC AAC ATC 1248
Ala Arg Asp Glu Lys Lys Lys Leu Arg His Ala Ser Gly Ala Asn Ile
                405                 410                 415
```

```
ACC AAC GCC ACC ACT GCC GCC ACC ACG GCC GCC ACT GCC ACC ACC ACT  1296
Thr Asn Ala Thr Thr Ala Ala Thr Thr Ala Ala Thr Ala Thr Thr Thr
            420                 425                 430

GCC AGC ACC GAG GGC AGC AAC AGT GAG AGT GAG GCC GAG AGC ACC GAG  1344
Ala Ser Thr Glu Gly Ser Asn Ser Glu Ser Glu Ala Glu Ser Thr Glu
        435                 440                 445

AAC AGC CCC ACC CCA TCG CCG CTG CAG AAG AAG GTC ACT GAG GAT CTG  1392
Asn Ser Pro Thr Pro Ser Pro Leu Gln Lys Lys Val Thr Glu Asp Leu
    450                 455                 460

TCC AAA ACC CTC CTG ATG TAC ACG GTG CCT GCC GTC CAG GGC TTC TTC  1440
Ser Lys Thr Leu Leu MET Tyr Thr Val Pro Ala Val Gln Gly Phe Phe
465                 470                 475                 480

CGT TCC ATC TCC TTG TCA CGA GGC AAC AAC CTC CAG GAT ACA CTC AGA  1488
Arg Ser Ile Ser Leu Ser Arg Gly Asn Asn Leu Gln Asp Thr Leu Arg
                485                 490                 495

GTT CTC ACC TTA TGG TTT GAT TAT GGT CAC TGG CCA GAT GTC AAT GAG  1536
Val Leu Thr Leu Trp Phe Asp Tyr Gly His Trp Pro Asp Val Asn Glu
            500                 505                 510

GCC TTA GTG GAG GGG GTG AAA GCC ATC CAG ATT GAT ACC TGG CTA CAG  1584
Ala Leu Val Glu Gly Val Lys Ala Ile Gln Ile Asp Thr Trp Leu Gln
        515                 520                 525

GTT ATA CCT CAG CTC ATT GCA AGA ATT GAT ACG CCC AGA CCC TTG GTG  1632
Val Ile Pro Gln Leu Ile Ala Arg Ile Asp Thr Pro Arg Pro Leu Val
    530                 535                 540

GGA CGT CTC ATT CAC CAG CTT CTC ACA GAC ATT GGT CGG TAC CAC CCC  1680
Gly Arg Leu Ile His Gln Leu Leu Thr Asp Ile Gly Arg Tyr His Pro
545                 550                 555                 560

CAG GCC CTC ATC TAC CCA CTG ACA GTG GCT TCT AAG TCT ACC ACG ACA  1728
Gln Ala Leu Ile Tyr Pro Leu Thr Val Ala Ser Lys Ser Thr Thr Thr
                565                 570                 575

GCC CGG CAC AAT GCA GCC AAC AAG ATT CTG AAG AAC ATG TGT GAG CAC  1776
Ala Arg His Asn Ala Ala Asn Lys Ile Leu Lys Asn MET Cys Glu His
            580                 585                 590

AGC AAC ACC CTG GTC CAG CAG GCC ATG ATG GTG AGC GAG GAG CTG ATC  1824
Ser Asn Thr Leu Val Gln Gln Ala MET MET Val Ser Glu Glu Leu Ile
        595                 600                 605

CGA GTG GCC ATC CTC TGG CAT GAG ATG TGG CAT GAA GGC CTG GAA GAG  1872
Arg Val Ala Ile Leu Trp His Glu MET Trp His Glu Gly Leu Glu Glu
    610                 615                 620
```

```
GCA TCT CGT TTG TAC TTT GGG GAA AGG AAC GTG AAA GGC ATG TTT GAG  1920
Ala Ser Arg Leu Tyr Phe Gly Glu Arg Asn Val Lys Gly MET Phe Glu
625                 630                 635                 640

GTG CTG GAG CCC TTG CAT GCT ATG ATG GAA CGG GGC CCC CAG ACT CTG  1968
Val Leu Glu Pro Leu His Ala MET MET Glu Arg Gly Pro Gln Thr Leu
                645                 650                 655

AAG GAA ACA TCC TTT AAT CAG GCC TAT GGT CGA GAT TTA ATG GAG GCC  2016
Lys Glu Thr Ser Phe Asn Gln Ala Tyr Gly Arg Asp Leu MET Glu Ala
            660                 665                 670

CAA GAG TGG TGC AGG AAG TAC ATG AAA TCA GGG AAT GTC AAG GAC CTC  2064
Gln Glu Trp Cys Arg Lys Tyr MET Lys Ser Gly Asn Val Lys Asp Leu
        675                 680                 685

ACC CAA GCC TGG GAC CTC TAT TAT CAT GTG TTC CGA CGA ATC TCA AAG  2112
Thr Gln Ala Trp Asp Leu Tyr Tyr His Val Phe Arg Arg Ile Ser Lys
    690                 695                 700

CAG CTG CCT CAG CTC ACA TCC TTA GAG CTG CAA TAT GTT TCC CCA AAA  2160
Gln Leu Pro Gln Leu Thr Ser Leu Glu Leu Gln Tyr Val Ser Pro Lys
705                 710                 715                 720

CTT CTG ATG TGC CGG GAC CTT GAA TTG GCT GTG CCA GGA ACA TAT GAC  2208
Leu Leu MET Cys Arg Asp Leu Glu Leu Ala Val Pro Gly Thr Tyr Asp
                725                 730                 735

CCC AAC CAG CCA ATC ATT CGC ATT CAG TCC ATA GCA CCG TCT TTG CAA  2256
Pro Asn Gln Pro Ile Ile Arg Ile Gln Ser Ile Ala Pro Ser Leu Gln
            740                 745                 750

GTC ATC ACA TCC AAG CAG AGG CCC CGG AAA TTG ACA CTT ATG GGC AGC  2304
Val Ile Thr Ser Lys Gln Arg Pro Arg Lys Leu Thr Leu MET Gly Ser
        755                 760                 765

AAC GGA CAT GAG TTT GTT TTC CTT CTA AAA GGC CAT GAA GAT CTG CGC  2352
Asn Gly His Glu Phe Val Phe Leu Leu Lys Gly His Glu Asp Leu Arg
    770                 775                 780

CAG GAT GAG CGT GTG ATG CAG CTC TTC GGC CTG GTT AAC ACC CTT CTG  2400
Gln Asp Glu Arg Val MET Gln Leu Phe Gly Leu Val Asn Thr Leu Leu
785                 790                 795                 800

GCC AAT GAC CCA ACA TCT CTT CGG AAA AAC CTC AGC ATC CAG AGA TAC  2448
Ala Asn Asp Pro Thr Ser Leu Arg Lys Asn Leu Ser Ile Gln Arg Tyr
                805                 810                 815

GCT GTC ATC CCT TTA TCG ACC AAC TCG GGC CTC ATT GGC TGG GTT CCC  2496
Ala Val Ile Pro Leu Ser Thr Asn Ser Gly Leu Ile Gly Trp Val Pro
            820                 825                 830
```

```
CAC TGT GAC ACA CTG CAC GCC CTC ATC CGG GAC TAC AGG GAG AAG AAG  2544
His Cys Asp Thr Leu His Ala Leu Ile Arg Asp Tyr Arg Glu Lys Lys
        835                 840                 845

AAG ATC CTT CTC AAC ATC GAG CAT CGC ATC ATG TTG CGG ATG GCT CCG  2592
Lys Ile Leu Leu Asn Ile Glu His Arg Ile MET Leu Arg MET Ala Pro
    850                 855                 860

GAC TAT GAC CAC TTG ACT CTG ATG CAG AAG GTG GAG GTG TTT GAG CAT  2640
Asp Tyr Asp His Leu Thr Leu MET Gln Lys Val Glu Val Phe Glu His
865                 870                 875                 880

GCC GTC AAT AAT ACA GCT GGG GAC GAC CTG GCC AAG CTG CTG TGG CTG  2688
Ala Val Asn Asn Thr Ala Gly Asp Asp Leu Ala Lys Leu Leu Trp Leu
                885                 890                 895

AAA AGC CCC AGC TCC GAG GTG TGG TTT GAC CGA AGA ACC AAT TAT ACC  2736
Lys Ser Pro Ser Ser Glu Val Trp Phe Asp Arg Arg Thr Asn Tyr Thr
            900                 905                 910

CGT TCT TTA GCG GTC ATG TCA ATG GTT GGG TAT ATT TTA GGC CTG GGA  2784
Arg Ser Leu Ala Val MET Ser MET Val Gly Tyr Ile Leu Gly Leu Gly
        915                 920                 925

GAT AGA CAC CCA TCC AAC CTG ATG CTG GAC CGT CTG AGT GGG AAG ATC  2832
Asp Arg His Pro Ser Asn Leu MET Leu Asp Arg Leu Ser Gly Lys Ile
    930                 935                 940

CTG CAC ATT GAC TTT GGG GAC TGC TTT GAG GTT GCT ATG ACC CGA GAG  2880
Leu His Ile Asp Phe Gly Asp Cys Phe Glu Val Ala MET Thr Arg Glu
945                 950                 955                 960

AAG TTT CCA GAG AAG ATT CCA TTT AGA CTA ACA AGA ATG TTG ACC AAT  2928
Lys Phe Pro Glu Lys Ile Pro Phe Arg Leu Thr Arg MET Leu Thr Asn
                965                 970                 975

GCT ATG GAG GTT ACA GGC CTG GAT GGC AAC TAC AGA ATC ACA TGC CAC  2976
Ala MET Glu Val Thr Gly Leu Asp Gly Asn Tyr Arg Ile Thr Cys His
            980                 985                 990

ACA GTG ATG GAG GTG CTG CGA GAG CAC AAG GAC AGT GTC ATG GCC GTG  3024
Thr Val MET Glu Val Leu Arg Glu His Lys Asp Ser Val MET Ala Val
        995                 1000                1005

CTG GAA GCC TTT GTC TAT GAC CCC TTG CTG AAC TGG AGG CTG ATG GAC  3072
Leu Glu Ala Phe Val Tyr Asp Pro Leu Leu Asn Trp Arg Leu MET Asp
    1010                1015                1020

ACA AAT ACC AAA GGC AAC AAG CGA TCC CGA ACG AGG ACG GAT TCC TAC  3120
Thr Asn Thr Lys Gly Asn Lys Arg Ser Arg Thr Arg Thr Asp Ser Tyr
1025                1030                1035                1040
```

```
TCT GCT GGC CAG TCA GTC GAA ATT TTG GAC GGT GTG GAA CTT GGA GAG  3168
Ser Ala Gly Gln Ser Val Glu Ile Leu Asp Gly Val Glu Leu Gly Glu
                1045                1050                1055

CCA GCC CAT AAG AAA ACG GGG ACC ACA GTG CCA GAA TCT ATT CAT TCT  3216
Pro Ala His Lys Lys Thr Gly Thr Thr Val Pro Glu Ser Ile His Ser
            1060                1065                1070

TTC ATT GGA GAC GGT TTG GTG AAA CCA GAG GCC CTA AAT AAG AAA GCT  3264
Phe Ile Gly Asp Gly Leu Val Lys Pro Glu Ala Leu Asn Lys Lys Ala
        1075                1080                1085

ATC CAG ATT ATT AAC AGG GTT CGA GAT AAG CTC ACT GGT CGG GAC TTC  3312
Ile Gln Ile Ile Asn Arg Val Arg Asp Lys Leu Thr Gly Arg Asp Phe
    1090                1095                1100

TCT CAT GAT GAC ACT TTG GAT GTT CCA ACG CAA GTT GAG CTG CTC ATC  3360
Ser His Asp Asp Thr Leu Asp Val Pro Thr Gln Val Glu Leu Leu Ile
1105                1110                1115                1120

AAA CAA GCG ACA TCC CAT GAA AAC CTC TGC CAG TGC TAT ATT GGC TGG  3408
Lys Gln Ala Thr Ser His Glu Asn Leu Cys Gln Cys Tyr Ile Gly Trp
                1125                1130                1135

TAC CCT TTC TGG TAA                                             3423
Tyr Pro Phe Trp
            1140
```